



PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : C12N 15/64, 15/70		A1	(11) International Publication Number: WO 92/06204		
(21) International Application Number: PCT/US91/07149		(81) Designated States: AT (European patent), AU, BB, BE (European patent), BF (OAPI patent), BG, BJ (OAPI patent), BR, CA, CF (OAPI patent), CG (OAPI patent), CH (European patent), CI (OAPI patent), CM (OAPI patent), CS, DE (European patent), DK (European patent), ES (European patent), FI, FR (European patent), GA (OAPI patent), GB (European patent), GN (OAPI patent), GR (European patent), HU, IT (European patent), JP, KP, KR, LK, LU (European patent), MC, MG, ML (OAPI patent), MN, MR (OAPI patent), MW, NL (European patent), NO, PL, RO, SD, SE (European patent), SN (OAPI patent), SU ⁺ , TD (OAPI patent), TG (OAPI patent).			
(22) International Filing Date: 27 September 1991 (27.09.91)		Published <i>With international search report.</i>			
(30) Priority data: 590,219 28 September 1990 (28.09.90) US					
(71) Applicant: IXSYS, INC. [US/US]; 3550 General Atomics Court, Suite L103, San Diego, CA 92121 (US).					
(72) Inventor: HUSE, William, D. ; 471 Avenida Primavera, Del Mar, CA 92014 (US).					
(74) Agents: CAMPBELL, Cathryn et al.; Pretty, Schroeder, Brueggemann & Clark, 444 South Flower Street, Suite 2000, Los Angeles, CA 90071 (US).					
(54) Title: SURFACE EXPRESSION LIBRARIES OF HETEROGENERIC RECEPTORS					
(57) Abstract					
A composition of matter comprising a plurality of prokaryotic cells containing diverse combinations of first and second DNA sequences encoding first and second polypeptides which form a heteromeric receptor exhibiting binding activity toward a preselected molecule, said heteromeric receptors being expressed on the surface of filamentous bacteriophage.					

⁺ See back of page

+ DESIGNATIONS OF "SU"

Any designation of "SU" has effect in the Russian Federation. It is not yet known whether any such designation has effect in other States of the former Soviet Union.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	ES	Spain	MC	Madagascar
AU	Australia	FI	Finland	ML	Mali
BB	Barbados	FR	France	MN	Mongolia
BE	Belgium	GA	Gabon	MR	Mauritania
BF	Burkina Faso	GB	United Kingdom	MW	Malawi
BG	Bulgaria	GN	Guinea	NL	Netherlands
BJ	Benin	GR	Greece	NO	Norway
BR	Brazil	HU	Hungary	PL	Poland
CA	Canada	IT	Italy	RO	Romania
CF	Central African Republic	JP	Japan	SD	Sudan
CG	Congo	KP	Democratic People's Republic of Korea	SE	Sweden
CH	Switzerland	KR	Republic of Korea	SN	Senegal
CI	Côte d'Ivoire	LI	Liechtenstein	SU+	Soviet Union
CM	Cameroon	LK	Sri Lanka	TD	Chad
CS	Czechoslovakia	LU	Luxembourg	TC	Togo
DE	Germany	MC	Monaco	US	United States of America
DK	Denmark				

SURFACE EXPRESSION LIBRARIES
OF HETEROGENERIC RECEPTORS

BACKGROUND OF THE INVENTION

This invention relates generally to recombinant
5 expression of heteromeric receptors and, more particularly,
to expression of such receptors on the surface of
filamentous bacteriophage.

Antibodies are heteromeric receptors generated by a
vertebrates organism's immune system which bind to an
10 antigen. The molecules are composed of two heavy and two
light chains disulfide bonded together. Antibodies have
the appearance of a "y" - shaped structure and the antigen
binding portion being located at the end of both short arms
of the y. The region on the heavy and light chain
15 polypeptides which corresponds to the antigen binding
portion is known as variable region. The differences
between antibodies within this region are primarily
responsible for the variation in binding specificities
between antibody molecules. The binding specificities are
20 a composite of the antigen interactions with both heavy and
light chain polypeptides.

The immune system has the capability of generating an
almost infinite number of different antibodies. Such a
large diversity is generated primarily through
25 recombination to form the variable regions of each chain
and through differential pairing of heavy and light chains.
The ability to mimic the natural immune system and generate
antibodies that bind to any desired molecule is valuable
because such antibodies can be used for diagnostic and
30 therapeutic purposes.

Until recently, generation of antibodies against a

desired molecule was accomplished only through manipulation of natural immune responses. Methods included classical immunization techniques of laboratory animals and monoclonal antibody production. Generation of monoclonal antibodies is laborious and time consuming. It involves a series of different techniques and is only performed on animal cells. Animal cells have relatively long generation times and require extra precautions to be taken compared to prokaryotic cells to ensure viability of the cultures.

A method for the generation of a large repertoire of diverse antibody molecules in bacteria has been described, Huse et al., Science, 246, 1275-1281 (1989), which is herein incorporated by reference. The method uses the bacteriophage lambda as the vector. The lambda vector is a long, linear double-stranded DNA molecule. Production of antibodies using this vector involves the cloning of heavy and light chain populations of DNA sequences into separate vectors. The vectors are subsequently combined randomly to form a single vector which directs the coexpression of heavy and light chains to form antibody fragments. A disadvantage to this method is that undesired combinations of vector portions are brought together when generating the coexpression vector. Although these undesired combinations do not produce viable phage, they do however, result in a significant loss of sequences from the population and, therefore, a loss in diversity of the number of different combinations which can be obtained between heavy and light chains. Additionally, the size of the lambda phage gene is large compared to the genes that encode the antibody segments. This makes the lambda system inherently more difficult to manipulate as compared to other available vector systems.

There thus exists a need for a method to generate diverse populations of heteromeric receptors which mimics the natural immune system, which is fast and efficient and

results in only desired combinations without loss of diversity. The present invention satisfies these needs and provides related advantages as well.

SUMMARY OF THE INVENTION

5 The invention relates to a plurality of cells containing diverse combinations of first and second DNA sequences encoding first and second polypeptides which form a heteromeric receptor, said heteromeric receptors being expressed on the surface of a cell, preferably one which
10 produces filamentous bacteriophage, such as M13. Vectors, cloning systems and methods of making and screening the heteromeric receptors are also provided.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a schematic diagram of the two vectors used for surface expression library construction from heavy and light chain libraries. M13IX30 (Figure 1A) is the vector used to clone the heavy chain sequences (open box). The single-headed arrow represents the Lac p/o expression sequences and the double-headed arrow represents the portion of M13IX30 which is to be combined with M13IX11. The amber stop codon and relevant restriction sites are also shown. M13IX11 (Figure 1B) is the vector used to clone the light chain sequences (hatched box). Thick lines represent the pseudo-wild type (gVIII) and wild type (gVIII) gene VIII sequences. The double-headed arrow represents the portion of M13IX11 which is to be combined with M13IX30. Relevant restriction sites are also shown. Figure 1C shows the joining of vector population from heavy and light chain libraries to form the functional surface expression vector M13IXHL. Figure 1D shows the generation of a surface expression library in a non-suppressor strain and the production of phage. The phage are used to infect a suppressor strain (Figure 1E) for surface expression and

screening of the library.

Figure 2 is the nucleotide sequence of M13IX30 (SEQ ID NO: 1).

Figure 3 is the nucleotide sequence of M13IX11 (SEQ ID NO: 2).

Figure 4 is the nucleotide sequence of M13IX34 (SEQ ID NO: 3).

Figure 5 is the nucleotide sequence of M13IX13 (SEQ ID NO: 4).

Figure 6 is the nucleotide sequence of M13IX60 (SEQ ID NO: 5).

DETAILED DESCRIPTION OF THE INVENTION

This invention is directed to simple and efficient methods to generate a large repertoire of diverse combinations of heteromeric receptors. The method is advantageous in that only proper combinations of vector portions are randomly brought together for the coexpression of different DNA sequences without loss of population size or diversity. The receptors can be expressed on the surface of cells, such as those producing filamentous bacteriophage, which can be screened in large numbers. The nucleic acid sequences encoding the receptors be readily characterized because the filamentous bacteriophage produce single strand DNA for efficient sequencing and mutagenesis methods. The heteromeric receptors so produced are useful in an unlimited number of diagnostic and therapeutic procedures.

In one embodiment, two populations of diverse heavy (Hc) and light (Lc) chain sequences are synthesized by

polymerase chain reaction (PCR). These populations are cloned into separate M13-based vector containing elements necessary for expression. The heavy chain vector contains a gene VIII (gVIII) coat protein sequence so that 5 translation of the Hc sequences produces gVIII-Hc fusion proteins. The populations of two vectors are randomly combined such that only the vector portions containing the Hc and Lc sequences are joined into a single circular vector. The combined vector directs the coexpression of 10 both Hc and Lc sequences for assembly of the two polypeptides and surface expression on M13. A mechanism also exists to control the expression of gVIII-Hc fusion proteins during library construction and screening.

As used herein, the term "heteromeric receptors" 15 refers to proteins composed of two or more subunits which together exhibit binding activity toward particular molecule. It is understood that the term includes the subunit fragments so long as assembly of the polypeptides and function of the assembled complex is retained. 20 Heteromeric subunits include, for example, antibodies and fragments thereof such as Fab and (Fab)₂ portions, T cell receptors, integrins, hormone receptors and transmitter receptors.

As used herein, the term "preselected molecule" refers 25 to a molecule which is chosen from a number of choices. The molecule can be, for example, a protein or peptide, or an organic molecule such as a drug. Benzodiazepam is a specific example of a preselected molecule.

As used herein, the term "coexpression" refers to the 30 expression of two or more nucleic acid sequences usually expressed as separate polypeptides. For heteromeric receptors, the coexpressed polypeptides assemble to form the heteromer. Therefore, "expression elements" as used herein, refers to sequences necessary for the

transcription, translation, regulation and sorting of the expressed polypeptides which make up the heteromeric receptors. The term also includes the expression of two subunit polypeptides which are linked but are able to 5 assemble into a heteromeric receptor. A specific example of coexpression of linked polypeptides is where Hc and Lc polypeptides are expressed with a flexible peptide or polypeptide linker joining the two subunits into a single chain. The linker is flexible enough to allow association 10 of Hc and Lc portions into a functional Fab fragment.

The invention provides for a composition of matter comprising a plurality of prokaryotic cells containing diverse combinations of first and second DNA sequences encoding first and second polypeptides which form a 15 heteromeric receptor exhibiting binding activity toward a preselected molecule, said heteromeric receptors being expressed on the surface of filamentous bacteriophage.

DNA sequences encoding the polypeptides of heteromeric receptors are obtained by methods known to one 20 skilled in the art. Such methods include, for example, cDNA synthesis and polymerase chain reaction (PCR). The need will determine which method or combinations of methods is to be used to obtain the desired populations of sequences. Expression can be performed in any compatible 25 vector/host system. Such systems include, for example, plasmids or phagemids in prokaryotes such as E. coli, yeast systems and other eucaryotic systems such as mammalian cells, but will be described herein in context with its presently preferred embodiment, i.e. expression on the 30 surface of filamentous bacteriophage. Filamentous bacteriophage include, for example, M13, f1 and fd. Additionally, the heteromeric receptors can also be expressed in soluble or secreted form depending on the need and the vector/host system employed.

Expression of heteromeric receptors such as antibodies or functional fragments thereof on the surface of M13 can be accomplished, for example, using the vector system shown in Figure 1. Construction of the vectors enabling one of ordinary skill to make them are explicitly set out in Example I. The complete nucleotide sequences are given in Figures 2 and 3 (SEQ ID NOS: 1 and 2). This system produces randomly combined populations of heavy (Hc) and light (Lc) chain antibody fragments functionally linked to expression elements. The Hc polypeptide is produced as a fusion protein with the M13 coat protein encoded by gene VIII. The gVIII-Hc fusion protein therefore anchors the assembled Hc and Lc polypeptides on the surface of M13. The diversity of Hc and Lc combinations obtained by this system can be 5×10^7 or greater. Diversity of less than 5×10^7 can also be obtained and will be determined by the need and type of heteromeric receptor to be expressed.

Populations of Hc and Lc encoding sequences to be combined into a vector for coexpression are each cloned into separate vectors. For the vectors shown in Figure 1, diverse populations of sequences encoding Hc polypeptides are cloned into M13IX30 (SEQ ID NO: 1). Sequences encoding Lc polypeptides are cloned into M13IX11 (SEQ ID NO: 2). The populations are inserted between the Xho I-Spe I or Stu I restriction enzyme sites in M13IX30 and between the Sac I-Xba I or Eco RV sites in M13IX11 (Figures 1A and B, respectively).

The populations of Hc and Lc sequences inserted into the vectors can be synthesized with appropriate restriction recognition sequences flanking opposite ends of the encoding sequences but this is not necessary. The sites allow annealing and ligation in-frame with expression elements of these sequences into a double-stranded vector restricted with the appropriate restriction enzyme. Alternatively, and a preferred embodiment, the Hc and Lc

sequences can be inserted into the vector without restriction of the DNA. This method of cloning is beneficial because naturally encoded restriction enzyme sites may be present within the sequences, thus, causing
5 destruction of the sequence when treated with a restriction enzyme. For cloning without restriction, the sequences are treated briefly with a 3' to 5' exonuclease such as T4 DNA polymerase or exonuclease III. A 5' to 3' exonuclease will also accomplish the same function. The protruding 5'
10 termini which remains should be complementary to single-stranded overhangs within the vector which remain after restriction at the cloning site and treatment with exonuclease. The exonuclease treated inserts are annealed with the restricted vector by methods known to one skilled
15 in the art. The exonuclease method decreases background and is easier to perform.

The vector used for Hc populations, M13IX30 (Figure 1A; SEQ ID NO: 1) contains, in addition to expression elements, a sequence encoding the pseudo-wild type gVIII product downstream and in frame with the cloning sites.
20 This gene encodes the wild type M13 gVIII amino acid sequence but has been changed at the nucleotide level to reduce homologous recombination with the wild type gVIII contained on the same vector. The wild type gVIII is
25 present to ensure that at least some functional, non-fusion coat protein will be produced. The inclusion of a wild type gVIII therefore reduces the possibility of non-viable phage production and biological selection against certain peptide fusion proteins. Differential regulation of the
30 two genes can also be used to control the relative ratio of the pseudo and wild type proteins.

Also contained downstream and in frame with the cloning sites is an amber stop codon. The stop codon is located between the inserted Hc sequences and the gVIII sequence and is in frame. As was the function of the wild

type gVIII, the amber stop codon also reduces biological selection when combining vector portions to produce functional surface expression vectors. This is accomplished by using a non-suppressor (sup 0) host strain 5 because the non-suppressor strains will terminate expression after the Hc sequences but before the pseudo gVIII sequences. Therefore, the pseudo gVIII will essentially never be expressed on the phage surface under these circumstances. Instead, only soluble Hc polypeptides 10 will be produced. Expression in a non-suppressor host strain can be advantageously utilized when one wishes to produce large populations of antibody fragments. Stop codons other than amber, such as opal and ochre, or molecular switches, such as inducible repressor elements, 15 can also be used to unlink peptide expression from surface expression.

The vector used for Lc populations, M13IX11 (SEQ ID NO: 2), contains necessary expression elements and cloning sites for the Lc sequences, Figure 1B. As with M13IX30, 20 upstream and in frame with the cloning sites is a leader sequence for sorting to the phage surface. Additionally, a ribosome binding site and Lac Z promoter/operator elements are also present for transcription and translation of the DNA sequences.

Both vectors contain two pairs of Mlu I-Hind III 25 restriction enzyme sites (Figures 1A and B) for joining together the Hc and Lc encoding sequences and their associated vector sequences. Mlu I and Hind III are non-compatible restriction sites. The two pairs are 30 symmetrically orientated about the cloning site so that only the vector portions containing the sequences to be expressed are exactly combined into a single vector. The two pairs of sites are oriented identically with respect to one another on both vectors and the DNA between the two 35 sites must be homologous enough between both vectors to

allow annealing. This orientation allows cleavage of each circular vector into two portions and combination of essential components within each vector into a single circular vector where the encoded polypeptides can be
5 coexpressed (Figure 1C).

Any two pairs of restriction enzyme sites can be used so long as they are symmetrically orientated about the cloning site and identically orientated on both vectors. The sites within each pair, however, should be non-
10 identical or able to be made differentially recognized as a cleavage substrate. For example, the two pairs of restriction sites contained within the vectors shown in Figure 1 are Mlu I and Hind III. The sites are differentially cleavable by Mlu I and Hind III
15 respectively. One skilled in the art knows how to substitute alternative pairs of restriction enzyme sites for the Mlu I-Hind III pairs described above. Also, instead of two Hind III and two Mlu I sites, a Hind III and Not I site can be paired with a Mlu I and a Sal I site, for
20 example.

The combining step randomly brings together different Hc and Lc encoding sequences within the two diverse populations into a single vector (Figure 1C; M13IXHL). The vector sequences donated from each independent vector,
25 M13IX30 and M13IX11, are necessary for production of viable phage. Also, since the pseudo gVIII sequences are contained in M13IX30, coexpression of functional antibody fragments as Lc associated gVIII-Hc fusion proteins cannot be accomplished on the phage surface until the vector
30 sequences are linked as shown in M13IXHL.

The combining step is performed by restricting each population of Hc and Lc containing vectors with Mlu I and Hind III, respectively. The 3' termini of each restricted vector population is digested with a 3' to 5' exonuclease

as described above for inserting sequences into the cloning sites. The vector populations are mixed, allowed to anneal and introduced into an appropriate host. A non-suppressor host (Figure 1D) is preferably used during initial 5 construction of the library to ensure that sequences are not selected against due to expression as fusion proteins. Phage isolated from the library constructed in a non-suppressor strain can be used to infect a suppressor strain for surface expression of antibody fragments.

10 A method for selecting a heteromeric receptor exhibiting binding activity toward a preselected molecule from a population of diverse heteromeric receptors, comprising: (a) operationally linking to a first vector a first population of diverse DNA sequences encoding a 15 diverse population of first polypeptides, said first vector having two pairs of restriction sites symmetrically oriented about a cloning site; (b) operationally linking to a second vector a second population of diverse DNA sequences encoding a diverse population of second 20 polypeptides, said second vector having two pairs of restriction sites symmetrically oriented about a cloning site in an identical orientation to that of the first vector; (c) combining the vector products of step (a) and (b) under conditions which allow only the operational 25 combination of vector sequences containing said first and second DNA sequences; (d) introducing said population of combined vectors into a compatible host under conditions sufficient for expressing said population of first and second DNA sequences; and (e) determining the heteromeric 30 receptors which bind to said preselected molecule. The invention also provides for determining the nucleic acid sequences encoding such polypeptides as well.

Surface expression of the antibody library is performed in an amber suppressor strain. As described 35 above, the amber stop codon between the Hc sequence and the

gVIII sequence unlinks the two components in a non-suppressor strain. Isolating the phage produced from the non-suppressor strain and infecting a suppressor strain will link the Hc sequences to the gVIII sequence during expression (Figure 1E). Culturing the suppressor strain after infection allows the coexpression on the surface of M13 of all antibody species within the library as gVIII fusion proteins (gVIII-Fab fusion proteins). Alternatively, the DNA can be isolated from the non-suppressor strain and then introduced into a suppressor strain to accomplish the same effect.

The level of expression of gVIII-Fab fusion proteins can additionally be controlled at the transcriptional level. Both polypeptides of the gVIII-Fab fusion proteins are under the inducible control of the Lac Z promoter/operator system. Other inducible promoters can work as well and are known by one skilled in the art. For high levels of surface expression, the suppressor library is cultured in an inducer of the Lac Z promoter such as isopropylthio- β -galactoside (IPTG). Inducible control is beneficial because biological selection against non-functional gVIII-Fab fusion proteins can be minimized by culturing the library under non-expressing conditions. Expression can then be induced only at the time of screening to ensure that the entire population of antibodies within the library are accurately represented on the phage surface. Also, this can be used to control the valency of the antibody on the phage surface.

The surface expression library is screened for specific Fab fragments which bind preselected molecules by standard affinity isolation procedures. Such methods include, for example, panning, affinity chromatography and solid phase blotting procedures. Panning as described by Parmley and Smith, Gene 73:305-318 (1988), which is incorporated herein by reference, is preferred because high

titors of phage can be screened easily, quickly and in small volumes. Furthermore, this procedure can select minor Fab fragments species within the population, which otherwise would have been undetectable, and amplified to 5 substantially homogenous populations. The selected Fab fragments can be characterized by sequencing the nucleic acids encoding the polypeptides after amplification of the phage population.

The following examples are intended to illustrate but 10 not limit the invention.

EXAMPLE I

Construction, Expression and Screening of Antibody Fragments on the Surface of M13

This example shows the synthesis of a diverse 15 population of heavy (Hc) and light (Lc) chain antibody fragments and their expression on the surface of M13 as gene VIII-Fab fusion proteins. The expressed antibodies derive from the random mixing and coexpression of a Hc and Lc pair. Also demonstrated is the isolation and 20 characterization of the expressed Fab fragments which bind benzodiazepam (BDP) and their corresponding nucleotide sequence.

Isolation of mRNA and PCR Amplification of Antibody Fragments

25 The surface expression library is constructed from mRNA isolated from a mouse that had been immunized with KLH-coupled benzodiazepam (BDP). BDP was coupled to keyhole limpet hemocyanin (KLH) using the techniques described in Antibodies: A Laboratory Manual, Harlow and 30 Lane, eds., Cold Spring Harbor, New York (1988), which is incorporated herein by reference. Briefly, 10.0 milligrams (mg) of keyhole limpet hemocyanin and 0.5 mg of BDP with a

glutaryl spacer arm N-hydroxysuccinimide linker appendages. Coupling was performed as in Jonda et al., Science, 241:1188 (1988), which is incorporated herein by reference. The KLH-BDP conjugate was removed by gel filtration 5 chromatography through Sephadex G-25.

The KLH-BDP conjugate was prepared for injection into mice by adding 100 µg of the conjugate to 250 µl of phosphate buffered saline (PBS). An equal volume of complete Freund's adjuvant was added and emulsified the 10 entire solution for 5 minutes. Mice were injected with 300 µl of the emulsion. Injections were given subcutaneously at several sites using a 21 gauge needle. A second immunization with BDP was given two weeks later. This injection was prepared as follows: 50 µg of BDP was 15 diluted in 250 µl of PBS and an equal volume of alum was mixed with the solution. The mice were injected intraperitoneally with 500 µl of the solution using a 23 gauge needle. One month later the mice were given a final injection of 50 µg of the conjugate diluted to 200 µl in 20 PBS. This injection was given intravenously in the lateral tail vein using a 30 gauge needle. Five days after this final injection the mice were sacrificed and total cellular RNA was isolated from their spleens.

Total RNA was isolated from the spleen of a single 25 mouse immunized as described above by the method of Chomczynski and Sacchi, Anal. Biochem., 162:156-159 (1987), which is incorporated herein by reference. Briefly, immediately after removing the spleen from the immunized mouse, the tissue was homogenized in 10 ml of a denaturing 30 solution containing 4.0 M guanine isothiocyanate, 0.25 M sodium citrate at pH 7.0, and 0.1 M 2-mercaptoethanol using a glass homogenizer. One ml of sodium acetate at a concentration of 2 M at pH 4.0 was mixed with the homogenized spleen. One ml of saturated phenol was also 35 mixed with the denaturing solution containing the

homogenized spleen. Two ml of a chloroform:isoamyl alcohol (24:1 v/v) mixture was added to this homogenate. The homogenate was mixed vigorously for ten seconds and maintained on ice for 15 minutes. The homogenate was then 5 transferred to a thick-walled 50 ml polypropylene centrifuge tube (Fisher Scientific Company, Pittsburgh, PA). The solution was centrifuged at 10,000 x g for 20 minutes at 4°C. The upper RNA-containing aqueous layer was transferred to a fresh 50 ml polypropylene centrifuge tube 10 and mixed with an equal volume of isopropyl alcohol. This solution was maintained at -20°C for at least one hour to precipitate the RNA. The solution containing the precipitated RNA was centrifuged at 10,000 x g for twenty minutes at 4°C. The pelleted total cellular RNA was 15 collected and dissolved in 3 ml of the denaturing solution described above. Three mls of isopropyl alcohol was added to the resuspended total cellular RNA and vigorously mixed. This solution was maintained at -20°C for at least 1 hour to precipitate the RNA. The solution containing the 20 precipitated RNA was centrifuged at 10,000 x g for ten minutes at 4°C. The pelleted RNA was washed once with a solution containing 75% ethanol. The pelleted R. . was dried under vacuum for 15 minutes and then resuspended in dimethyl pyrocarbonate (DEPC) treated (DEPC-H₂O) H₂O.

25 Poly A⁺ RNA for use in first strand cDNA synthesis was prepared from the above isolated total RNA using a spin-column kit (Pharmacia, Piscataway, NJ) as recommended by the manufacturer. The basic methodology has been described by Aviv and Leder, Proc. Natl. Acad. Sci., USA, 69:1408-30 1412 (1972), which is incorporated herein by reference. Briefly, one half of the total RNA isolated from a single immunized mouse spleen prepared as described above was resuspended in one ml of DEPC-treated dH₂O and maintained at 35 65°C for five minutes. One ml of 2x high salt loading buffer (100 mM Tris-HCL at pH 7.5, 1 M sodium chloride, 2.0 mM disodium ethylene diamine tetraacetic acid (EDTA) at pH

8.0, and 0.2% sodium dodecyl sulfate (SDS)) was added to the resuspended RNA and the mixture was allowed to cool to room temperature. The mixture was then applied to an oligo-dT (Collaborative Research Type 2 or Type 3 Bedford, MA) column that was previously prepared by washing the oligo-dT with a solution containing 0.1 M sodium hydroxide and 5 mM EDTA and then equilibrating the column with DEPC-treated dH₂O. The eluate was collected in a sterile polypropylene tube and reapplied to the same column after heating the eluate for 5 minutes at 65°C. The oligo dT column was then washed with 2 ml of high salt loading buffer consisting of 50 mM Tris-HCL at pH 7.5, 500 mM sodium chloride, 1 mM EDTA at pH 8.0 and 0.1% SDS. The oligo dT column was then washed with 2 ml of 1 X medium salt buffer (50 mM Tris-HCL at pH 7.5, 100 mM sodium chloride, 1 mM EDTA at pH 8.0 and 0.1% SDS). The mRNA was eluted with 1 ml of buffer consisting of 10 mM Tris-HCL at pH 7.5, 1 mM EDTA at pH 8.0 and 0.05% SDS. The messenger RNA was purified by extracting this solution with phenol/chloroform followed by a single extraction with 100% chloroform, ethanol precipitated and resuspended in DEPC treated dH₂O.

In preparation for PCR amplification, mRNA was used as a template for cDNA synthesis. In a typical 250 µl reverse transcription reaction mixture, 5-10 µg of spleen mRNA in water was first annealed with 500 ng (0.5 pmol) of either the 3' V_H primer (primer 12, Table I) or the 3' V_L primer (primer 9, Table II) at 65°C for 5 minutes. Subsequently, the mixture was adjusted to contain 0.8 mM dATP, 0.8 mM dCTP, 0.8 mM dGTP, 0.8 mM dTTP, 100 mM Tris-HCL (pH 8.6), 10 mM MgCl₂, 40 mM KCl, and 20 mM 2-ME. Moloney-Murine Leukemia Virus (Bethesda Research Laboratories (BRL), Gaithersburg, MD) Reverse transcriptase, 26 units, was added and the solution was incubated for 1 hour at 40°C. The resultant first strand cDNA was phenol extracted, ethanol precipitated and then used in the polymerase chain

reaction (PCR) procedures described below for amplification of heavy and light chain sequences.

Primers used for amplification of heavy chain Fd fragments for construction of the M13IX30 library is shown 5 in Table I. Amplification was performed in eight separate reactions, as described by Saiki et al., Science, 239:487-491 (1988), which is incorporated herein by reference, each reaction containing one of the 5' primers (primers 2 to 9; SEQ ID NOS: 7 through 14, respectively) and one of the 3' 10 primers (primer 12; SEQ ID NO: 17) listed in Table I. The remaining 5' primers, used for amplification in a single reaction, are either a degenerate primer (primer 1; SEQ ID NO: 6) or a primer that incorporates inosine at four degenerate positions (primer 10; SEQ ID NO: 15). The 15 remaining 3' primer (primer 11; SEQ ID NO: 16) was used to construct Fv fragments. The underlined portion of the 5' primers incorporates an Xho I site and that of the 3' primer an Spe I restriction site for cloning the amplified fragments into the M13IX30 vector in a predetermined 20 reading frame for expression.

TABLE I
HEAVY CHAIN PRIMERS

		CC G G T
25	1)	5' - AGGT A CT <u>CTCGAGTC</u> GG - 3' GA A T A
	2)	5' - AGGTCCAGCTG <u>CTCGAGT</u> CTGG - 3'
	3)	5' - AGGTCCAGCT <u>GCTCGAGT</u> CAGG - 3'
	4)	5' - AGGTCCAG <u>CTTCTCGAGT</u> CTGG - 3'
	5)	5' - AGGTCCAG <u>CTTCTCGAGT</u> CAGG - 3'
30	6)	5' - AGGTCCA <u>ACTGCTCGAGT</u> CTGG - 3'
	7)	5' - AGGTCCA <u>ACTGCTCGAGT</u> CAGG - 3'
	8)	5' - AGGTCCA <u>ACTTCTCGAGT</u> CTGG - 3'

- 9) 5' - AGGTCCAACTTTCGAGTCAGG - 3'
- 10) 5' - AGGTIIIAICTITCGAGTC ^T
A GG - 3'
- 5 11) 5' - CTATTAACTAGTAACGGTAACAGT -
GGTGCCTTGCCCA - 3'
- 12) 5' - AGGCTTACTAGTACAATCCCTGG -
GCACAAT - 3'

Primers used for amplification of mouse kappa light chain sequences for construction of the M13IX11 library are shown in Table II. These primers were chosen to contain restriction sites which were compatible with vector and not present in the conserved sequences of the mouse light chain mRNA. Amplification was performed as described above in five separate reactions, each containing one of the 5' primers (primers 3 to 7; SEQ ID NOS: 20 through 24, respectively) and one of the 3' primers (primer 9; SEQ ID NO: 26) listed in Table II. The remaining 3' primer (primer 8; SEQ ID NO: 25) was used to construct Fv fragments. The underlined portion of the 5' primers depicts a Sac I restriction site and that of the 3' primers an Xba I restriction site for cloning of the amplified fragments into the M13IX11 vector in a predetermined reading frame for expression.

25

TABLE II
LIGHT CHAIN PRIMERS

- 1) 5' - CCAGTTCCGAGCTCGTGACTCAGGAATCT - 3'
2) 5' - CCAGTTCCGAGCTCGTGTTGACGCAGCCGCC - 3'
3) 5' - CCAGTTCCGAGCTCGTGCTCACCCAGTCTCCA - 3'
30 4) 5' - CCAGTTCCGAGCTCCAGATGACCCAGTCTCCA - 3'
5) 5' - CCAGATGTGAGCTCGTGATGACCCAGACTCCA - 3'
6) 5' - CCAGATGTGAGCTCGTCATGACCCAGTCTCCA - 3'
7) 5' - CCAGTTCCGAGCTCGTGATGACACAGTCTCCA - 3'
8) 5' - GCAGCATTCTAGAGTTCAGCTCCAGCTTGCC - 3'
35 9) 5' - GCGCCGTCTAGAATTAAACACTATTCCGTGAA - 3'

PCR amplification for heavy and light chain fragments was performed in a 100 μ l reaction mixture containing the above described products of the reverse transcription reaction (\approx 5 μ g of the cDNA-RNA hybrid), 300 nmol of 3' V_H primer (primer 12, Table I; SEQ ID NO: 17), and one of the 5' V_H primers (primers 2-9, Table I; SEQ ID NOS: 7 through 14, respectively) for heavy chain amplification, or, 300 nmol of 3' V_L primer (primer 9, Table III; SEQ ID NO: 26), and one of the 5' V_L primers (primers 3-7, Table II; SEQ ID NOS: 20 through 24, respectively) for each light chain amplification, a mixture of dNTPs at 200 mM, 50 mM KCl, 10 mM Tris-HCl (pH 8.3), 15 mM MgCl₂, 0.1% gelatin, and 2 units of *Thermus aquaticus* DNA polymerase. The reaction mixture was overlaid with mineral oil and subjected to 40 cycles of amplification. Each amplification cycle involved denaturation at 92°C for 1 minute, annealing at 52°C for 2 minutes, and elongation at 72°C for 1.5 minutes. The amplified samples were extracted twice with phenol/CHCl₃, and once with CHCl₃, ethanol-precipitated, and stored at -70°C in 10 mM Tris-HCl, pH 7.5 1 mM EDTA. The resultant products were used in constructing the M13IX30 and M13IX11 libraries (see below).

Vector Construction

Two M13-based vectors, M13IX30 (SEQ ID NO: 1) and M13IX11 (SEQ ID NO: 2), were constructed for the cloning and propagation of Hc and Lc populations of antibody fragments, respectively. The vectors were constructed to facilitate the random joining and subsequent surface expression of antibody fragment populations.

M13IX30 (SEQ ID NO: 1), or the Hc vector, was constructed to harbor diverse populations of Hc antibody fragments. M13mp19 (Pharmacia, Piscataway, NJ) was the starting vector. This vector was modified to contain, in addition to the encoded wild type M13 gene VIII: (1) a

pseudo-wild type gene VIII sequence with an amber stop codon between it and the restriction sites for cloning oligonucleotides; (2) Stu I restriction site for insertion of sequences by hybridization and, Spe I and Xho I 5 restriction sites in-frame with the pseudo-wild type gene VIII for cloning Hc sequences; (3) sequences necessary for expression, such as a promoter, signal sequence and translation initiation signals; (4) two pairs of Hind III-Mlu I sites for random joining of Hc and Lc vector 10 portions, and (5) various other mutations to remove redundant restriction sites and the amino terminal portion of Lac Z.

Construction of M13IX30 was performed in four steps. In the first step, an M13-based vector containing the 15 pseudo gVIII and various other mutations was constructed, M13IX01F. The second step involved the construction of a small cloning site in a separate M13mp18 vector to yield M13IX03. This vector was then expanded to contain expression sequences and restriction sites for Hc sequences 20 to form M13IX04B. The fourth and final step involved the incorporation of the newly constructed sequences in M13IX04B into M13IX01F to yield M13IX30.

Construction of M13IX01F first involved the generation of a pseudo wild-type gVIII sequence for surface expression 25 of antibody fragments. The pseudo-wild type gene encodes the identical amino acid sequence as that of the wild type gene; however, the nucleotide sequence has been altered so that only 63% identity exists between this gene and the encoded wild type gene VIII. Modification of the gene VIII 30 nucleotide sequence used for surface expression reduces the possibility of homologous recombination with the wild type gene VIII contained on the same vector. Additionally, the wild type M13 gene VIII was retained in the vector system to ensure that at least some functional, non-fusion coat 35 protein would be produced. The inclusion of wild type gene

VIII facilitates the growth of phage under conditions where there is surface expression of the polypeptides and therefore reduces the possibility of non-viable phage production from the fusion genes.

- 5 The pseudo-wild type gene VIII was constructed by chemically synthesizing a series of oligonucleotides which encode both strands of the gene. The oligonucleotides are presented in Table III.

TABLE IIIPseudo-Wild Type Gene VIII Oligonucleotide Series

	<u>Top Strand Oligonucleotides</u>	<u>Sequence (5' to 3')</u>
5	VIII 03	GATCC TAG GCT GAA GGC GAT GAC CCT GCT AAG GCT GC
	VIII 04	A TTC AAT AGT TTA CAG GCA AGT GCT ACT GAG TAC A
10	VIII 05	TT GGC TAC GCT TGG GCT ATG GTA GTA GTT ATA GTT GGT GCT ACC ATA GGG ATT
	VIII 06	AAA TTA TTC AAA AAG TT
15	VIII 07	T ACG AGC AAG GCT TCT TA
	<u>Bottom Strand Oligonucleotides</u>	
20	VIII 08	AGC TTA AGA AGC CTT GCT CGT AAA CTT TTT GAA TAA TTT
	VIII 09	AAT CCC TAT GGT AGC ACC AAC TAT AAC TAC TAC CAT
	VIII 10	AGC CCA AGC GTA GCC AAT GTA CTC AGT AGC ACT TG
25	VIII 11	C CTG TAA ACT ATT GAA TGC AGC CTT AGC AGG GTC
	VIII 12	ATC GCC TTC AGC CTA G

Except for the terminal oligonucleotides VIII 03 (SEQ ID NO: 27) and VIII 08 (SEQ ID NO: 32), the above oligonucleotides (oligonucleotides VIII 04-07 (SEQ ID NOS: 28 through 31, respectively) and VIII 09-12 (SEQ ID NOS: 33

through 36, respectively)) were mixed at 200 ng each in 10 μ l final volume, phosphorylated with T4 polynucleotide Kinase (Pharmacia) and 1 mM ATP at 37°C for 1 hour, heated to 70°C for 5 minutes, and annealed into double-stranded 5 form by heating to 65°C for 3 minutes, followed by cooling to room temperature over a period of 30 minutes. The reactions were treated with 1.0 U of T4 DNA ligase (BRL) and 1 mM ATP at room temperature for 1 hour, followed by heating to 70°C for 5 minutes. Terminal oligonucleotides 10 were then annealed to the ligated oligonucleotides. The annealed and ligated oligonucleotides yielded a double-stranded DNA flanked by a Bam HI site at its 5' end and by a Hind III site at its 3' end. A translational stop codon (amber) immediately follows the Bam HI site. The gene VIII 15 sequence begins with the codon GAA (Glu) two codons 3' to the stop codon. The double-stranded insert was cloned in frame with the Eco RI and Sac I sites within the M13 polylinker. To do so, M13mp19 was digested with Bam HI (New England Biolabs, Beverley, MA) and Hind III (New 20 England Biolabs) and combined at a molar ratio of 1:10 with the double-stranded insert. The ligations were performed at room temperature overnight in 1X ligase buffer (50 mM Tris-HCl, pH 7.8, 10 mM MgCl₂, 20 mM DTT, 1 mM ATP, 50 μ g/ml BSA) containing 1.0 U of T4 DNA ligase (New England 25 Biolabs). The ligation mixture was transformed into a host and screened for positive clones using standard procedures in the art.

Several mutations were generated within the construct to yield functional M13IX01F. The mutations were generated 30 using the method of Kunkel et al., Meth. Enzymol. 154:367-382 (1987), which is incorporated herein by reference, for site-directed mutagenesis. The reagents, strains and protocols were obtained from a Bio Rad Mutagenesis kit (Bio Rad, Richmond, CA) and mutagenesis was performed as 35 recommended by the manufacturer.

Two Fok I sites were removed from the vector as well as the Hind III site at the end of the pseudo gene VIII sequence using the mutant oligonucleotides 5'-CATTGGCAGATGGCTTAGA-3' (SEQ ID NO: 37) and 5'-
5 TAGCATTAACGTCCAATA-3' (SEQ ID NO: 38). New Hind III and Mlu I sites were also introduced at position 3919 and 3951 of M13IX01F. The oligonucleotides used for this mutagenesis had the sequences 5'-ATATATTTAGTAAGCTTCATCTTCT-3' (SEQ ID NO: 39) and 5'-
10 GACAAAGAACGCGTGAAAACCTT-3' (SEQ ID NO: 40), respectively. The amino terminal portion of Lac Z was deleted by oligonucleotide-directed mutagenesis using the mutant oligonucleotide 5'-GCAGGGCCTCTCGCTATTGCTTAAGAACGCTTGCT-3' (SEQ ID NO: 41). In constructing the above mutations, all
15 changes made in a M13 coding region were performed such that the amino acid sequence remained unaltered. The resultant vector, M13IX01F, was used in the final step to construct M13IX30 (see below).

In the second step, M13mp18 was mutated to remove the
20 5' end of Lac Z up to the Lac i binding site and including the Lac Z ribosome binding site and start codon. Additionally, the polylinker was removed and a Mlu I site was introduced in the coding region of Lac Z. A single oligonucleotide was used for these mutagenesis and had the
25 sequence 5'-AAACGACGCCAGTGCCAAAGTGACGCGTGTGAAATTGTTATCC-3' (SEQ ID NO: 42). Restriction enzyme sites for Hind III and Eco RI were introduced downstream of the Mlu I site using the oligonucleotide 5'-GGCGAAAGGAATTCTGCAAGGCGATTAAGCTTGGG
TAACGCC-3' (SEQ ID NO. 43). These modifications of M13mp18
30 yielded the precursor vector M13IX03.

The expression sequences and cloning sites were introduced into M13IX03 by chemically synthesizing a series of oligonucleotides which encode both strands of the desired sequence. The oligonucleotides are presented in
35 Table IV.

TABLE IV
M13IX30 Oligonucleotide Series

<u>Top Strand Oligonucleotides</u>		<u>Sequence (5' to 3')</u>
5	084	GGCGTTACCCAAGCTTGATGGAGAAAATAAAG
	027	TGAAACAAAGCACTATTGCACTGGCACTCTTACCGT TACCGT
	028	TACTGTTACCCCTGTGACAAAAGCCGCCAGGTCC AGCTGC
10	029	TCGAGTCAGGCCTATTGTGCCAGGGATTGTACTAG TGGATCCG
<u>Bottom Oligonucleotides</u>		<u>Sequence (5' to 3')</u>
	085	TGGCGAAAGGAATTGGATCCACTAGTACAATCCCTG
15	031	GGCACAAATAGGCCTGACTCGAGCAGCTGGACCAGGGCG GCTT
	032	TTGTCACAGGGTAAACAGTAACGGTAACGGTAAGTGT GCCA
20	033	GTGCAATAGTGCTTGTTCACTTTATTTCTCCATGT ACAA

The above oligonucleotides of Table IV, except for the terminal oligonucleotides 084 (SEQ ID NO: 44) and 085 (SEQ ID NO: 48), were mixed, phosphorylated, annealed and ligated to form a double-stranded insert as described in Example I. However, instead of cloning directly into the intermediate vector the insert was first amplified by PCR. The terminal oligonucleotides were used as primers for PCR. Oligonucleotide 084 (SEQ ID NO: 44) contains a Hind III site, 10 nucleotides internal to its 5' end and oligonucleotide 085 (SEQ ID NO: 48) has an Eco RI site at its 5' end. Following amplification, the products were restricted with Hind III and Eco RI and ligated, as described in Example I, into the polylinker of M13mp18 digested with the same two enzymes. The resultant double

stranded insert contained a ribosome binding site, a translation initiation codon followed by a leader sequence and three restriction enzyme sites for cloning random oligonucleotides (Xho I, Stu I, Spe I). The intermediate 5 vector was named M13IX04.

During cloning of the double-stranded insert, it was found that one of the GCC codons in oligonucleotides 028 and its complement in 031 was deleted. Since this deletion did not affect function, the final construct is missing one 10 of the two GCC codons. Additionally, oligonucleotide 032 (SEQ ID NO: 50) contained a GTG codon where a GAG codon was needed. Mutagenesis was performed using the oligonucleotide 5'-TAACGGTAAGAGTGCCAGTGC-3' (SEQ ID NO: 52) to convert the codon to the desired sequence. The 15 resultant vector is named M13IX04B.

The third step in constructing M13IX30 involved inserting the expression and cloning sequences from M13IX04B upstream of the pseudo wild-type gVIII in M13IX01F. This was accomplished by digesting M13IX04B with 20 Dra III and Bam HI and gel isolating the 700 base pair insert containing the sequences of interest. M13IX01F was likewise digested with Dra III and Bam HI. The insert was combined with the double digested vector at a molar ratio of 1:1 and ligated as described in Example I. The sequence 25 of the final construct M13IX30, is shown in Figure 2 (SEQ ID NO: 1). Figure 1A also shows M13IX30 where each of the elements necessary for surface expression of Hc fragments is marked. It should be noted during modification of the vectors, certain sequences differed from the published 30 sequence of M13mp18. The new sequences are incorporated into the sequences recorded herein.

M13IX11 (SEQ ID NO: 2), or the Lc vector, was constructed to harbor diverse populations of Lc antibody fragments. This vector was also constructed from M13mp19

and contains: (1) sequences necessary for expression, such as a promoter, signal sequence and translation initiation signals; (2) Eco RV restriction site for insertion of sequences by hybridization and Sac I and Xba I restriction sites for cloning of Lc sequences; (3) two pairs of Hind III-Mlu I sites for random joining of Hc and Lc vector portions, and (4) various other mutation to remove redundant restriction sites.

The expression, translation initiation signals, cloning sites, and one of the Mlu I sites were constructed by annealing of overlapping oligonucleotides as described above to produce a double-stranded insert containing a 5' Eco RI site and a 3' Hind III site. The overlapping oligonucleotides are shown in Table V and were ligated as a double-stranded insert between the Eco RI and Hind III sites of M13mp18 as described for the expression sequences inserted into M13IX03. The ribosome binding site (AGGAGAC) is located in oligonucleotide 015 and the translation initiation codon (ATG) is the first three nucleotides of oligonucleotide 016 (SEQ ID NO: 55).

TABLE V

Oligonucleotide Series for Construction of
Translation Signals in M13IX11

	<u>Oligonucleotide</u>	<u>Sequence (5' to 3')</u>
5	082	CACC TTCATG AATTC GGC AAG GAGACA GTCAT
	015	AATT C GCC AAG GAG ACA GTC AT
	016	AATG AAA TAC CTA TTG CCT ACG GCA GCC GCT GGA TTG TT
10	017	ATTA CTC GCT GCC CAA CCA GCC ATG GCC GAG CTC GTG AT
	018	GACC CAG ACT CCA GATATC CAA CAG GAA TGA GTG TTA AT
	019	TCT AGA ACG CGT C
15	083	TTCAGGTTGAAGC TTA CGC GTT CTA GAA TTA ACA CTC ATT
	021	CCTGT TG GAT ATC TGG AGT CTG GGT CAT CAC GAG CTC GGC CAT G
20	022	GC TGG TTG GGC AGC GAG TAA TAA CAA TCC AGC GGC TGC C
	023	GT AGG CAA TAG GTA TTT CAT TAT GAC TGT CCT TGG CG

Oligonucleotide 017 (SEQ ID NO: 56) contained a Sac I restriction site 67 nucleotides downstream from the ATG codon. The naturally occurring Eco RI site was removed and new Eco RI and Hind III sites were introduced downstream from the Sac I. Oligonucleotides 5'-TGACTGTCTCCTTGGCGTGTGAAATTGTTA-3' (SEQ ID NO: 63) and 5'-TAACACTCATTCCGGATGGAATTCTGGAGTCTGGGT-3' (SEQ ID NO: 64) were used to generate each of the mutations, respectively. The Lac Z ribosome binding site was removed when the

original Eco RI site in M13mp19 was mutated. Additionally, when the new Eco RI and Hind III sites were generated, a spontaneous 100 bp deletion was found just 3' to these sites. Since the deletion does not affect the function, it
5 was retained in the final vector.

In addition to the above mutations, a variety of other modifications were made to incorporate or remove certain sequences. The Hind III site used to ligate the double-stranded insert was removed with the oligonucleotide 5'-
10 GCCAGTGCCAAAGTGACGCGTTCTA-3' (SEQ ID NO: 65). Second Hind III and Mlu I sites were introduced at positions 3922 and 3952, respectively, using the oligonucleotides 5'-ATATATTAGTAAGCTTCATCTTCT-3' (SEQ ID NO: 66) for the Hind III mutagenesis and 5'-GACAAAGAACCGCGTGAAAACCTT-3' (SEQ ID
15 NO: 67) for the Mlu I mutagenesis. Again, mutations within the coding region did not alter the amino acid sequence.

The sequence of the resultant vector, M13IX11, is shown in Figure 3 (SEQ ID NO: 2). Figure 1B also shows M13IX11 where each of the elements necessary for producing
20 a surface expression library between Lc fragments is marked.

Library Construction

Each population of Hc and Lc sequences synthesized by PCR above are separately cloned into M13IX30 and M13IX11,
25 respectively, to create Hc and Lc libraries.

The Hc and Lc products (5 µg) are mixed, ethanol precipitated and resuspended in 20 µl of NaOAc buffer (33 mM Tris acetate, pH 7.9, 10 mM Mg-acetate, 66 mM K-acetate, 0.5 mM DTT). Five units of T4 DNA polymerase is added and
30 the reactions incubated at 30°C for 5 minutes to remove 3' termini by exonuclease digestion. Reactions are stopped by heating at 70°C for 5 minutes. M13IX30 is digested with

Stu I and M13IX11 is digested with Eco RV. Both vectors are treated with T4 DNA polymerase as described above and combined with the appropriate PCR products at a 1:1 molar ratio at 10 ng/ μ l to anneal in the above buffer at room 5 temperature overnight. DNA from each annealing is electroporated into MK30-3 (Boehringer, Indianapolis, IN), as described below, to generate the Hc and Lc libraries.

E. coli MK30-3 is electroporated as described by Smith et al., Focus 12:38-40 (1990) which is incorporated herein 10 by reference. The cells are prepared by inoculating a fresh colony of MK30-3 into 5 mls of SOB without magnesium (20 g bacto-tryptone, 5 g bacto-yeast extract, 0.584 g NaCl, 0.186 g KC1, dH₂O to 1,000 mls) and grown with vigorous aeration overnight at 37°C. SOB without magnesium 15 (500 ml) is inoculated at 1:1000 with the overnight culture and grown with vigorous aeration at 37°C until the OD₅₅₀ is 0.8 (about 2 to 3 h). The cells are harvested by centrifugation at 5,000 rpm (2,600 x g) in a GS3 rotor (Sorvall, Newtown, CT) at 4°C for 10 minutes, resuspended 20 in 500 ml of ice-cold 10% (v/v) sterile glycerol, centrifuged and resuspended a second time in the same manner. After a third centrifugation, the cells are resuspended in 10% sterile glycerol at a final volume of about 2 ml, such that the OD₅₅₀ of the suspension was 200 to 25 300. Usually, resuspension is achieved in the 10% glycerol that remained in the bottle after pouring off the supernate. Cells are frozen in 40 μ l aliquots in microcentrifuge tubes using a dry ice-ethanol bath and stored frozen at -70°C.

30 Frozen cells are electroporated by thawing slowly on ice before use and mixing with about 10 pg to 500 ng of vector per 40 μ l of cell suspension. A 40 μ l aliquot is placed in an 0.1 cm electroporation chamber (Bio-Rad, Richmond, CA) and pulsed once at 0°C using 4 k Ω parallel 35 resistor 25 μ F, 1.88 KV, which gives a pulse length (τ) of

~4 ms. A 10 μ l aliquot of the pulsed cells are diluted into 1 ml SOC (98 mls SOB plus 1 ml of 2 M MgCl₂ and 1 ml of 2 M glucose) in a 12- x 75-mm culture tube, and the culture is shaken at 37°C for 1 hour prior to culturing in 5 selective media, (see below).

Each of the libraries are cultured using methods known to one skilled in the art. Such methods can be found in Sanbrook et al., Molecular Cloning: A Laboratory Manuel, Cold Spring Harbor Laboratory, Cold Spring Harbor, 1989, 10 and in Ausubel et al., Current Protocols in Molecular Biology, John Wiley and Sons, New York, 1989, both of which are incorporated herein by reference. Briefly, the above 1 ml library cultures are grown up by diluting 50-fold into 2XYT media (16 g tryptone, 10 g yeast extract, 5 g NaCl) 15 and culturing at 37°C for 5-8 hours. The bacteria are pelleted by centrifugation at 10,000 \times g. The supernatant containing phage is transferred to a sterile tube and stored at 4°C.

Double strand vector DNA containing Hc and Lc antibody 20 fragments are isolated from the cell pellet of each library. Briefly, the pellet is washed in TE (10 mM Tris, pH 8.0, 1 mM EDTA) and recollected by centrifugation at 7,000 rpm for 5' in a Sorval centrifuge (Newtown, CT). Pellets are resuspended in 6 mls of 10% Sucrose, 50 mM 25 Tris, pH 8.0. 3.0 ml of 10 mg/ μ l lysozyme is added and incubated on ice for 20 minutes. 12 mls of 0.2 M NaOH, 1% SDS is added followed by 10 minutes on ice. The suspensions are then incubated on ice for 20 minutes after addition of 7.5 mls of 3 M NaOAc, pH 4.6. The samples are 30 centrifuged at 15,000 rpm for 15 minutes at 4°C, RNased and extracted with phenol/chloroform, followed by ethanol precipitation. The pellets are resuspended, weighed and an equal weight of CsCl₂ is dissolved into each tube until a density of 1.60 g/ml is achieved. EtBr is added to 600 35 μ g/ml and the double-stranded DNA is isolated by

equilibrium centrifugation in a TV-1665 rotor (Sorval) at 50,000 rpm for 6 hours. These DNAs from each right and left half sublibrary are used to generate forty libraries in which the right and left halves of the randomized 5 oligonucleotides have been randomly joined together.

The surface expression library is formed by the random joining of the Hc containing portion of M13IX30 with the Lc containing portion of M13IX11. The DNAs isolated from each library was digested separately with an excess amount of 10 restriction enzyme. The Lc population (5 µg) is digested with Hind III. The Hc (5 µg) population is digested with Mlu I. The reactions are stopped by phenol/chloroform extraction followed by ethanol precipitation. The pellets are washed in 70% ethanol and resuspended in 20 µl of NaOAc 15 buffer. Five units of T4 DNA polymerase (Pharmacia) is added and the reactions incubated at 30°C for 5 minutes. Reactions are stopped by heating at 70°C for 5 minutes. The Hc and Lc DNAs are mixed to a final concentration of 10 ng each vector/µl and allowed to anneal at room temperature 20 overnight. The mixture is electroporated into MK30-3 cells as described above.

Screening of Surface Expression Libraries

Purified phage are prepared from 50 ml liquid cultures of XL1 Blue™ cells (Stratagene, La Jolla, CA) which had 25 been infected at a m.o.i. of 10 from the phage stocks stored at 4°C. The cultures are induced with 2 mM IPTG. Supernatants are cleared by two centrifugations, and the phage are precipitated by adding 1/7.5 volumes of PEG solution (25% PEG-8000, 2.5 M NaCl), followed by incubation 30 at 4°C overnight. The precipitate is recovered by centrifugation for 90 minutes at 10,000 x g. Phage pellets are resuspended in 25 ml of 0.01 M Tris-HCl, pH 7.6, 1.0 mM EDTA, and 0.1% Sarkosyl and then shaken slowly at room temperature for 30 minutes. The solutions are adjusted to

- 0.5 M NaCl and to a final concentration of 5% polyethylene glycol. After 2 hours at 4°C, the precipitates containing the phage are recovered by centrifugation for 1 hour at 15,000 X g. The precipitates are resuspended in 10 ml of
- 5 NET buffer (0.1 M NaCl, 1.0 mM EDTA, and 0.01 M Tris-HCl, pH 7.6), mixed well, and the phage repelleted by centrifugation at 170,000 X g for 3 hours. The phage pellets are resuspended overnight in 2 ml of NET buffer and subjected to cesium chloride centrifugation for 18 hours at
- 10 110,000 X g (3.86 g of cesium chloride in 10 ml of buffer). Phage bands are collected, diluted 7-fold with NET buffer, re-centrifuged at 170,000 X g for 3 hours, resuspended, and stored at 4°C in 0.3 ml of NET buffer containing 0.1 mM sodium azide.
- 15 The BDP used for panning on streptavidin coated dishes is first biotinylated and then absorbed against UV-inactivated blocking phage (see below). The biotinylation reagents are dissolved in dimethylformamide at a ratio of 2.4 mg solid NHS-SS-Biotin (sulfosuccinimidyl 2-
- 20 (biotinamido)ethyl-1,3'-dithiopropionate; Pierce, Rockford, IL) to 1 ml solvent and used as recommended by the manufacturer. Small-scale reactions are accomplished by mixing 1 µl dissolved reagent with 43 µl of 1 mg/ml BDP diluted in sterile bicarbonate buffer (0.1 M NaHCO₃, pH
- 25 8.6). After 2 hours at 25°C, residual biotinylation reagent is reacted with 500 µl 1 M ethanolamine (pH adjusted to 9 with HCl) for an additional 2 hours. The entire sample is diluted with 1 ml TBS containing 1 mg/ml BSA, concentrated to about 50 µl on a Centricon 30 ultra-
- 30 filter (Amicon), and washed on the same filter three times with 2 ml TBS and once with 1 ml TBS containing 0.02% NaN₃ and 7 x 10¹² UV-inactivated blocking phage (see below); the final retentate (60-80 µl) is stored at 4 °C. BDP biotinylated with the NHS-SS-Biotin reagent is linked to
- 35 biotin via a disulfide-containing chain.

UV-irradiated M13 phage are used for blocking any biotinylated BDP which fortuitously binds filamentous phage in general. M13mp8 (Messing and Vieira, Gene 19: 262-276 (1982), which is incorporated herein by reference) is 5 chosen because it carries two amber mutations, which ensure that the few phage surviving irradiation will not grow in the sup O strains used to titer the surface expression library. A 5 ml sample containing 5×10^{13} M13mp8 phage, purified as described above, is placed in a small petri 10 plate and irradiated with a germicidal lamp at a distance of two feet for 7 minutes (flux 150 $\mu\text{W}/\text{cm}^2$). NaN_3 is added to 0.02% and phage particles concentrated to 10^{14} particles/ml on a Centricon 30-kDa ultrafilter (Amicon).

For panning, polystyrene petri plates (60 x 15 mm) are 15 incubated with 1 ml of 1 mg/ml of streptavidin (BRL) in 0.1 M NaHCO_3 pH 8.6-0.02% NaN_3 , in a small, air-tight plastic box overnight in a cold room. The next day streptavidin is removed and replaced with at least 10 ml blocking solution (29 mg/ml of BSA; 3 $\mu\text{g}/\text{ml}$ of streptavidin; 0.1 M NaHCO_3 pH 20 8.6-0.02% NaN_3) and incubated at least 1 hour at room temperature. The blocking solution is removed and plates are washed rapidly three times with Tris buffered saline containing 0.5% Tween 20 (TBS-0.5% Tween 20).

Selection of phage expressing antibody fragments which 25 bind BDP is performed with 5 μl (2.7 μg BDP) of blocked biotinylated BDP reacted with a 50 μl portion of the library. Each mixture is incubated overnight at 4°C, diluted with 1 ml TBS-0.5% Tween 20, and transferred to a streptavidin-coated petri plate prepared as described 30 above. After rocking 10 minutes at room temperature, unbound phage are removed and plates washed ten times with TBS-0.5% Tween 20 over a period of 30-90 minutes. Bound phage are eluted from plates with 800 μl sterile elution buffer (1 mg/ml BSA, 0.1 M HCl, pH adjusted to 2.2 with 35 glycerol) for 15 minutes and eluates neutralized with 48 μl

2 M Tris (pH unadjusted). A 20 μ l portion of each eluate is titered on MK30-3 concentrated cells with dilutions of input phage.

A second round of panning is performed by treating 750
5 μ l of first eluate from the library with 5 mM DTT for 10 minutes to break disulfide bonds linking biotin groups to residual biotinylated binding proteins. The treated eluate is concentrated on a Centricon 30 ultrafilter (Amicon),
washed three times with TBS-0.5% Tween 20, and concentrated
10 to a final volume of about 50 μ l. Final retentate is transferred to a tube containing 5.0 μ l (2.7 μ g BDP) blocked biotinylated BDP and incubated overnight. The solution is diluted with 1 ml TBS-0.5% Tween 20, panned,
and eluted as described above on fresh streptavidin-coated
15 petri plates. The entire second eluate (800 μ l) is neutralized with 48 μ l 2 M Tris, and 20 μ l is titered simultaneously with the first eluate and dilutions of the input phage. If necessary, further rounds of panning can be performed to obtain homogeneous populations of phage.
20 Additionally, phage can be plaque purified if reagents are available for detection.

Template Preparation and Sequencing

Templates are prepared for sequencing by inoculating a 1 ml culture of 2XYT containing a 1:100 dilution of an
25 overnight culture of XL1 with an individual plaque from the purified population. The plaques are picked using a sterile toothpick. The culture is incubated at 37°C for 5-6 hours with shaking and then transferred to a 1.5 ml microfuge tube. 200 μ l of PEG solution is added, followed
30 by vortexing and placed on ice for 10 minutes. The phage precipitate is recovered by centrifugation in a microfuge at 12,000 x g for 5 minutes. The supernatant is discarded and the pellet is resuspended in 230 μ l of TE (10 mM Tris-HCl, pH 7.5, 1 mM EDTA) by gently pipeting with a yellow

pipet tip. Phenol (200 μ l) is added, followed by a brief vortex and microfuged to separate the phases. The aqueous phase is transferred to a separate tube and extracted with 200 μ l of phenol/chloroform (1:1) as described above for
5 the phenol extraction. A 0.1 volume of 3 M NaOAc is added, followed by addition of 2.5 volumes of ethanol and precipitated at -20°C for 20 minutes. The precipitated templates are recovered by centrifugation in a microfuge at 12,000 \times g for 8 minutes. The pellet is washed in 70%
10 ethanol, dried and resuspended in 25 μ l TE. Sequencing was performed using a Sequenase™ sequencing kit following the protocol supplied by the manufacturer (U.S. Biochemical, Cleveland, OH).

EXAMPLE II

15 Cloning of Heavy and Light Chain Sequences
 Without Restriction Enzyme Digestion

This example shows the simultaneous incorporation of antibody heavy and light chain fragment encoding sequences into a M13IXHL-type vector with the use of restriction
20 endonucleases.

For the simultaneous incorporation of heavy and light chain encoding sequences into a single coexpression vector, a M13IXHL vector was produced that contained heavy and light chain encoding sequences for a mouse monoclonal
25 antibody (DAN-18H4; Biosite, San Diego, CA). The inserted antibody fragment sequences are used as complementary sequences for the hybridization and incorporation of Hc and Lc sequences by site-directed mutagenesis. The genes encoding the heavy and light chain polypeptides were
30 inserted into M13IX30 (SEQ ID NO: 1) and M13IX11 (SEQ ID NO: 2), respectively, and combined into a single surface expression vector as described in Example I. The resultant M13IXHL-type vector is termed M13IX50.

The combinations were performed under conditions that facilitate the formation of one Hc and one Lc vector half into a single circularized vector. Briefly, the overhangs generated between the pairs of restriction sites after 5 restriction with Mlu I or Hind III and exonuclease digestion are unequal (i.e., 64 nucleotides compared to 32 nucleotides). These unequal lengths result in differential hybridization temperatures for specific annealing of the complementary ends from each vector. The specific 10 hybridization of each end of each vector half was accomplished by first annealing at 65°C in a small volume (about 100 µg/µl) to form a dimer of one Hc vector half and one Lc vector half. The dimers were circularized by diluting the mixture (to about 20 µg/µl) and lowering the 15 temperature to about 25-37°C to allow annealing. T4 ligase was present to covalently close the circular vectors.

M13IX50 was modified such that it did not produce a functional polypeptide for the DAN monoclonal antibody. To do this, about eight amino acids were changed within the 20 variable region of each chain by mutagenesis. The Lc variable region was mutagenized using the oligonucleotide 5'-CTAACCTGTCAGGGACCACAGTTGATGCTATAGGATCAGATCTAGAATTCAATTAGAGACTGGCCTGGCTTCTGC-3' (SEQ ID NO: 68). The Hc sequence was mutagenized with the oligonucleotide 5'-
25 T C G A C C C G T T G G T A G G A A T A A A T G C A A T T A A T G GAGTAGCTCAAATTCAAGAACAGTACACCCAGTGCATCCAGTAGCT-3' (SEQ ID NO: 69). An additional mutation was also introduced into M13IX50 to yield the final form of the vector. During construction of an intermediate to M13IX50 (M13IX04
30 described in Example I), a six nucleotide sequence was duplicated in oligonucleotide 027 and its complement 032. This sequence, 5'TTACCG-3' was deleted by mutagenesis using the oligonucleotide 5'-GGTAAACAGTAACGGTAAGAGTGCCAG-3' (SEQ ID NO: 70). The resultant vector was designated M13IX53.

35 M13IX53 can be produced as a single stranded form and

contains all the functional elements of the previously described M13IXHL vector except that it does not express functional antibody heteromers. The single-stranded vector can be hybridized to populations of single-stranded Hc and
5 Lc encoding sequences for their incorporation into the vector by mutagenesis. Populations of single-stranded Hc and Lc encoding sequences can be produced by one skilled in the art from the PCR products described in Example I or by other methods known to one skilled in the art using the
10 primers and teachings described therein. The resultant vectors with Hc and Lc encoding sequences randomly incorporated are propagated and screened for desired binding specificities as described in Example I.

Other vectors similar to M13IX53 and the vectors it's
15 derived from, M13IX11 and M13IX30, have also been produced for the incorporation of Hc and Lc encoding sequences without restriction. In contrast to M13IX53, these vectors contain human antibody sequences for the efficient hybridization and incorporation of populations of human Hc
20 and Lc sequences. These vectors are briefly described below. The starting vectors were either the Hc vector (M13IX30) or the Lc vector (M13IX11) previously described.

M13IX32 was generated from M13IX30 by removing the six nucleotide redundant sequence 5'-TTACCG-3' described above
25 and mutation of the leader sequence to increase secretion of the product. The oligonucleotide used to remove the redundant sequence is the same as that given above. The mutation in the leader sequence was generated using the oligonucleotide 5'GGGCTTTGCCACAGGGT-3'. This mutagenesis
30 resulted in the A residue at position 6353 of M13IX30 being changed to a G residue.

A decapeptide tag for affinity purification of antibody fragments was incorporated in the proper reading frame at the carboxy-terminal end of the Hc expression site

in M13IX32. The oligonucleotide used for this mutagenesis was 5'-CGCCTT CAGCCTAAGAACGCTAGTCCGGAACGTCGTACGGTAGGATCCA CTAG-3' (SEQ ID NO: 71). The resultant vector was designated M13IX33. Modifications to this or other vectors 5 are envisioned which include various features known to one skilled in the art. For example, a peptidase cleavage site can be incorporated following the decapeptide tag which allows the antibody to be cleaved from the gene VIII portion of the fusion protein.

10 M13IX34 (SEQ ID NO: 3) was created from M13IX33 by cloning in the gene encoding a human IgG1 heavy chain. The reading frame of the variable region was changed and a stop codon was introduced to ensure that a functional polypeptide would not be produced. The oligonucleotide 15 used for the mutagenesis of the variable region was 5'-CACCGGTT CGGGGAATTAGTCTTGACCAGGCAGCCCAGGGC-3' (SEQ ID NO: 72). The complete nucleotide sequence of this vector is shown in Figure 4 (SEQ ID NO: 3).

Several vectors of the M13IX11 series were also 20 generated to contain similar modifications as that described for the vectors M13IX53 and M13IX34. The promoter region in M13IX11 was mutated to conform to the 35 consensus sequence to generate M13IX12. The oligonucleotide used for this mutagenesis was 5'-ATTCCACAC 25 ATTATACCGAGCCGGAAAGCATAAAGTGTCAAGCCTGGGTGCC-3' (SEQ ID NO: 73). A human kappa light chain sequence was cloned into M13IX12 and the variable region subsequently deleted to generate M13IX13 (SEQ ID NO: 4). The complete nucleotide sequence of this vector is shown in Figure 5 (SEQ ID NO: 30 4). A similar vector, designated M13IX14, was also generated in which the human lambda light chain was inserted into M13IX12 followed by deletion of the variable region. The oligonucleotides used for the variable region deletion of M13IX13 and M13IX14 were 5'-CTG 35 CTCATCAGATGGCGGGAAAGAGCTCGGCCATGGCTGGTTG-3' (SEQ ID NO: 74)

and 5'-GAACAGAGT GACCGAGGGGGCGAGCTCGGCCATGGCTGGTTG-3' (SEQ ID NO: 75), respectively.

The Hc and Lc vectors or modified forms thereof can be combined using the methods described in Example I to produce a single vector similar to M13IX53 that allows the efficient incorporation of human Hc and Lc encoding sequences by mutagenesis. An example of such a vector is the combination of M13IX13 with M13IX34. The complete nucleotide sequence of this vector, M13IX60, is shown in Figure 6 (SEQ ID NO: 5).

Additional modifications to any of the previously described vectors can also be performed to generate vectors which allow the efficient incorporation and surface expression of Hc and Lc sequences. For example, to alleviate the use of uracil selection against wild-type template during mutagenesis procedures, the variable region locations within the vectors can be substituted by a set of palindromic restriction enzyme sites (i.e., two similar sites in opposite orientation). The palindromic sites will loop out and hybridize together during the mutagenesis and thus form a double-stranded substrate for restriction endonuclease digestion. Cleavage of the site results in the destruction of the wild-type template. The variable region of the inserted Hc or Lc sequences will not be affected since they will be in single stranded form.

Following the methods of Example I, single-stranded Hc or Lc populations can be produced by a variety of methods known to one skilled in the art. For example, the PCR primers described in Example I can be used in asymmetric PCR to generate such populations. Gelfand et al., "PCR Protocols: A Guide to Methods and Applications", Ed by M.A. Innis (1990), which is incorporated herein by reference. Asymmetric PCR is a PCR method that differentially amplifies only a single strand of the double

stranded template. Such differential amplification is accomplished by decreasing the primer amount for the undesirable strand about 10-fold compared to that for the desirable strand. Alternatively, single-stranded 5 populations can be produced from double-stranded PCR products generated as described in Example I except that the primer(s) used to generate the undesirable strand of the double-stranded products is first phosphorylated at its 5' end with a kinase. The resultant products can then be 10 treated with a 5' to 3' exonuclease, such as lambda exonuclease (BRL, Bethesda, MD) to digest away the unwanted strand.

Single-stranded Hc and Lc populations generated by the methods described above or by others known to one skilled 15 in the art are hybridized to complementary sequences encoded in the previously described vectors. The population of the sequences are subsequently incorporated into a double-stranded form of the vector by polymerase extension of the hybridized templates. Propagation and 20 surface expression of the randomly combined Hc and Lc sequences are performed as described in Example I.

Although the invention has been described with reference to the presently preferred embodiment, it should be understood that various modifications can be made 25 without departing from the spirit of the invention. Accordingly, the invention is limited only by the claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: HUSE, WILLIAM D.
- (ii) TITLE OF INVENTION: SURFACE EXPRESSION LIBRARIES OF HETEROMERIC RECEPTORS
- (iii) NUMBER OF SEQUENCES: 75
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: PRETTY, SCHROEDER, BRUEGEMANN & CLARK
 - (B) STREET: 444 SO. FLOWER STREET, SUITE 200
 - (C) CITY: LOS ANGELES
 - (D) STATE: CALIFORNIA
 - (E) COUNTRY: UNITED STATES
 - (F) ZIP: 90071
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.25
- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER:
 - (B) FILING DATE:
 - (C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: CAMPBELL, CATHRYN A.
 - (B) REGISTRATION NUMBER: 31,815
 - (C) REFERENCE/DOCKET NUMBER: P31 8882
- (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: 619-535-9001
 - (B) TELEFAX: 619-535-8949

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 7445 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: both
 - (D) TOPOLOGY: circular

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

AATGCTACTA CTATTAGTAG AATTGATGCC ACCTTTTCAG CTCGGCCCC AAATGAAAAT	60
ATAGCTAAC AGGTTATTGA CCATTTGCGA AATGTATCTA ATGGTCAAAC TAAATCTACT	120
CGTTCCGAGA ATTGGGAATC AACTGTTACA TGGAATGAAA CTTCCAGACA CCGTACTTTA	180
GTTGCATATT TAAAACATGT TGAGCTACAG CACCAGATTC AGCAATTAAG CTCTAAGCCA	240
TCTGCAAAAA TGACCTCTTA TCAAAAGGAG CAATTAAAGG TACTCTCTAA TCCTGACCTG	300
TTGGAGTTTG CTTCCGGTCT GGTTGGCTTT GAAGCTCGAA TTAAAACGCG ATATTTGAAG	360
TCTTTCGGGC TTCCTCTTAA TCTTTTGAT GCAATCCGCT TTGCTTCTGA CTATAATAGT	420

CACCGCTAAAG ACCTGATTT TGATTTATGG TCATTCTGGT TTTCTGAAC	480
TTTGAGGGGG ATTCAATGAA TATTTATGAC GATTCCGCAG TATTGGACGC TATCCAGTCT	540
AAACATTTA CTATTACCCC CTCTGGCAAA ACTTCCTTTG CAAAAGCCTC TCGCTATTT	600
GGTTTTATC GTCGCTGGT AAACGAGGGT TATGATAGTG TTGCTTAC TATGCCGT	660
AATTCCCTTT GGCGTTATGT ATCTGCATTA GTTGAATGTG GTATTCCCAA ATCTCAACTG	720
ATGAATCTT CTACCTGTAA TAATGTTGTT CCGTTAGTTC GTTTTATTAA CGTAGATTT	780
TCTTCCCAAC GTCCTGACTG GTATAATGAG CCAGTTCTTA AAATCGCATA AGGTAATTCA	840
CAATGATTAA AGTTGAAATT AAACCATCTC AAGCCCCATT TACTACTCGT TCTGGTGT	900
CTCGTCAGGG CAAGCCTTAT TCAC TGAAATG AGCAGCTTG TTACGTTGAT TTGGGTAATG	960
AATATCCGGT TCTTGTCAAG ATTACTCTTG ATGAAGGTCA GCCAGCCTAT GCGCCTGGTC	1020
TGTACACCGT TCATCTGTCC TCTTCAAAG TTGGTCAGTT CGGTTCCCTT ATGATTGACC	1080
GTCTGCCCT CGTTCCGGCT AAGTAACATG GAGCAGGTGG CGGATTCGA CACAATTTAT	1140
CAGGCGATGA TACAAATCTC CGTTGTACTT TGTTTCCGCG TTGGTATAAT CGCTGGGGT	1200
CAAAGATGAG TGTTTAGTG TATTCCTTCG CCTCTTCTGT TTTAGGTTGG TGCCCTCGTA	1260
GTGGCATTAC GTATTTAAC CGTTAATGG AAACCTCCTC ATGAAAAAGT CTTIAGTCCT	1320
CAAAGCCTCT GTAGCCGTG CTACCCCTCGT TCCGATGCTG TCTTCGCTG CTGAGGGTGA	1380
CGATCCCCA AAAGCCGCT TAAACTCCCT GCAAGCCTCA CGCACCGAAT ATATCGGTTA	1440
TGGTGGGGCG ATGGTTGTG TCATTGTCGG CGCAACTATC CGTATCAAGC TGTTTAAGAA	1500
ATTACCI.JG AAAGCAAGCT GATAAACCGA TACAATTAAA GGCTCCTTT GGAGCCTTT	1560
TTTTGGAGA TTTTCAACGT GAAAAAATTA TTATTGCAA TTCCCTTAGT TGTTCCCTTC	1620
TATTCTCACT CCGCTGAAAC TGTTGAAAGT TGTTAGCAA AACCCCATAC AGAAAATTCA	1680
TTTACTAACCG TCTGGAAAGA CGACAAAAGT TTAGATCGTT ACCTAACTA TGAGGGTTGT	1740
CTGTGGAATG CTACAGCGT TGTAGTTGT ACTGGTGACG AAACCTCAGTG TTACGGTACA	1800
TGGGTTCCCTA TTGGGCTTGC TATCCCTGAA AATGAGGGTG GTGGCTCTGA GGGTGGCGGT	1860
TCTGAGGGTG CGGGTTCTGA GGGTGGCGGT ACTAACCTC CTGAGTACGG TGATACACCT	1920
ATTCCGGGCT ATACTTATAT CAACCCCTCTC GACGGCACTT ATCCGCTGG TACTGAGCAA	1980
AAACCCGCTA ATCTAACTCG TTCTCTTGAG GAGTCTCGC CTCTTAATAC TTTCATGTTT	2040
CAGAATAATA GTTCCGAAA TAGGCAGGGG GCATTAACGT TTTATACGGG CACTGTTACT	2100
CAAGGCACGT ACCCCGTTAA AACTTATTAC CAGTACACTC CTGTATCATC AAAAGCCATG	2160
TATGACGCTT ACTGGAACGG TAAATTGAGA GACTGGCCTT TCCATTCTGG CTTTAATGAA	2220
GATCCATTG TTTGTGAATA TCAAGGCCAA TCGTCTGACG TGCCTCAACC TCCGTCAAT	2280
GCTGGCGGGCG GCTCTGGTGG TGTTCTGGT GCGGGCTCTG AGGGTGGTGG CTCTGAGGGT	2340
GGCGGTTCTG AGGGTGGCGG CTCTGAGGGGA GGCGGTTCCG GTGGTGGCTC TGGTTCCGGT	2400
GATTTTGATT ATGAAAAGAT GGCAAACGCT AATAAGGGGG CTATGACCGA AAATGCCGAT	2460

GAAAACGGCG TACAGTCTGA CGCTAAAGGC AAACTTGATT CTGTCGCTAC TGATTACGGT	2520
GCTGCTATCG ATGGTTTCAT TGGTGACGTT TCCGGCCTTG CTAATGGTAA TGGTGCTACT	2580
GGTGATTTTG CTGGCTCTAA TTCCCAAATG GCTCAAGTCG GTGACGGTGA TAATTCACCT	2640
TIAATGAATA ATTTCCGTCA ATATTTACCT TCCCTCCCTC AATCGGTTGA ATGTCGCCCT	2700
TTTGTCTTTA GCGCTGGTAA ACCATATGAA TTTTCTATTG ATTGTGACAA AATAAACTTA	2760
TTCCGTGGTG TCITTCGCGTT TCTTTATAT GTTGCCACCT TTATGTATGT ATTTTCTACG	2820
TTTGCTAACAA TACTGGTAA TAAGGAGTCT TAATCATGCC AGTTCTTTG GGTATTCCGT	2880
TATTATTGGG TTTCCTCGGT TTCCCTCTGG TAACTTGTT CGGCTATCTG CTTACTTTTC	2940
TTAAAAAGGG CTTCCGTAAAG ATAGCTATTG CTATTCATT GTTTCCTGCT CTTATTATTG	3000
GGCTTAACTC AATTCTTGTG GGTTATCTCT CTGATATTAG CGCTCAATTAA CCCCTGACT	3060
TIGTTCAAGGG TGTTCAAGTTA ATTCTCCCGT CTAATGCCGT TCCCTGTTT TATGTTATTG	3120
TCTCTGTAAA GGCTGCTATT TTCATTTTG ACGTTAAACA AAAAATCGTT TCTTATTG	3180
ATTGGGATAAA ATAATATGGC TGTTTATTGTT GTAACTGGCA AATTAGGCTG TGAAAAGACG	3240
CTCGTTAGCG TTGGTAAGAT TCAGGATAAA ATTGTAGCTG GGTGCAAAT AGCAACTAAT	3300
CTTGATTTAA GGCTTCAAAA CCTCCCGCAA GTCGGGAGGT TCGCTAAAC GCCTCGCGTT	3360
CTTAGAATAC CGGATAAGCC TTCTATATCT GATTTGCTTG CTATTGGCG CGGTAATGAT	3420
TCCTACGATG AAAATAAAA CGGCTTGCTT GTTCTCGATG AGTGGGCTAC TTGGTTAAAT	3480
ACCCGTTCTT GGAATGATAA GGAAAGACAG CCGATTATTG ATTGGTTCT ACATGCTCGT	3540
AAATTAGGAT GGGATATTAT TTTCCTGTT CAGGACTTAT CTATTGTTGA TAAACAGGGG	3600
CGTTCTGCAT TAGCTGAACA TGTTGTTTAT TGTCGTCGTC TGGACAGAAAT TACTTACCT	3660
TTTGTGGTA CTTTATATTG TCTTATTACT GGCTCGAAAA TGCCCTGCC TAAATTACAT	3720
GTGGCGTTG TTAAATATGG CGATTCTCAA TTAAAGCCCTA CTGTTGAGCG TTGGTTTAT	3780
ACTGGTAAGA ATTTGTATAA CGCATATGAT ACTAAACAGG CTTTTCTAG TAATTATGAT	3840
TCCGGTGTATTT ATTCTTATTT AACGCCCTAT TTATCACAGG GTCGGTATTT CAAACCATTAA	3900
AATTAGGTC AGAAGATGAA GCTTACTAAA ATATATTGA AAAAGTTTC ACGCGTTCTT	3960
TGTCTTGCGA TTGGATTTGC ATCAGCATT ACATATAGTT ATATAACCCA ACCTAACCGG	4020
GAGGTTAAAA AGGTAGTCTC TGAGACCTAT GATTTGATA AATTCACTAT TGACTCTCT	4080
CAGCGTCTTA ATCTAAGCTA TCGCTATGTT TTCAAGGATT CTAAGGGAAA ATTAATTAAAT	4140
AGCGACGATT TACAGAAGCA AGGTTATTCA CTCACATATA TTGATTATG TACTGTTCC	4200
ATTAAGGAAAG GTAATTCAA TGAAATTGTT AAATGTAATT AATTGTTTT TCTTGATGTT	4260
TGTTTCATCA TCTTCTTTG CTCAGGTAAT TGAAATGAAT AATTCCGCTC TGCGCGATT	4320
TGTAACCTGG TATTCAAAGC AATCAGGCCA ATCCGTTATT GTTTCTCCCG ATGTAAGG	4380
TACTGTTACT GTATATTCACT CTGACGTTAA ACCTGAAAAT CTACGCAATT TCTTTATTTC	4440
TGTTTACGT GCTAATAATT TTGATATGGT TGGTTCAATT CCTTCCATAA TTCAGAAGTA	4500

45

TAATCCAAAC AATCAGGATT ATATTGATGA ATTGCCATCA TCTGATAATC AGGAATATGA	4560
TGATAATTCC GCTCCCTCTG GTGGTTCTT TGTTCCCAA AATGATAATG TTACTCAAAC	4620
TTTTAAAATT AATAACGTTG GGGCAAAGGA TTTAATACGA GTTGTGAAT TGTTTGAAA	4680
GTCTAATACT TCTAAATCCT CAAATGTATT ATCTATTGAC GGCTCTAATC TATTAGTTGT	4740
TAATGCCACCT AAAGATATTT TAGATAACCT TCCTCAATTG CTTTCTACTG TTGATTTGCC	4800
AACTGACCGAG ATATTGATTG AGGGTTGAT ATTTGAGGTT CAGCAAGGTG ATGCTTTAGA	4860
TTTTTCATTT GCTGCTGGCT CTCAGCGTGG CACTGTTGCA GGCGGTGTTA ATACTGACCG	4920
CCTCACCTCT GTTTTATCTT CTGCTGGTGG TTCGTTCGGT ATTTTAATG GCGATGTTT	4980
AGGGCTATCA GTTCGGCAT TAAAGACTAA TAGGCATTCA AAAATATTGT CTGTGCCACGG	5040
TATTCTTACG CTTTCAGGTC AGAAGGGTTC TATCTCTGTT GGCCAGAATG TCCCTTTAT	5100
TAATGGTCGT GTGACTGGTG AATCTGCCAA TGAAATAAT CCATTTCAGA CGATTGAGCG	5160
TCAAAATGTA GGTATTTCCA TGAGCGTTT TCCTGTTGCA ATGGCTGGCG GTAATATTGT	5220
TCTGGATATT ACCAGCAAGG CCGATAGTTT GAGTTCTTCT ACTCAGGCAA GTGATGTTAT	5280
TACTAATCAA AGAAAGTATTG CTACAAACGGT TAATTTGGGT GATGGACAGA CTCTTTIAC	5340
CGGTGGCCTC ACTGATTATA AAAACACTTC TCAAGATTCT GGCGTACCGT TCCGTCTAA	5400
AATCCCTTTA ATCGGCCCTCC TGTTTAGCTC CCGCTCTGAT TCCAACGAGG AAAGCACGTT	5460
ATACGTGCTC GTCAAAGCAA CCATAGTACG CGCCCTGTAG CGGGCGATTA AGCGCGGGGG	5520
GTGTGGTGGT TACGGCGAGC GTGACCGCTA CACTTGCCAG CGCCCTAGGG CCCGCTCCTT	5580
TCCCTTTCTT CCCCTCCTTT CTGCCACGT TCGCCGGCTT TCCCCGTCAA GCTCTAAATC	5640
GGGGGCTCCC TTIAGGGTTC CGATTTAGTG CTTTACGGCA CCTCGACCCC AAAAAACTTG	5700
ATTTGGGTGA TGGTTCACGT AGTGGGCCAT CGCCCTGATA GACGGTTTTT CGCCCTTGA	5760
CGTTGGAGTC CACGTTCTTT AATAGTGGAC TCTTGTCCA AACTGGAACA ACACCTCAACC	5820
CTATCTCGGG CTATTCTTTT GATTATAAG GGATTTGCC GATTTGGAA CCACCATCAA	5880
ACAGGATTTT CGCCTGCTGG GGCIAACCGAG CGTGGACCGC TTGCTGCAAC TCTCTCAGGG	5940
CCAGGCGGTG AAGGGCAATC AGCTGTTGCC CGTCTCGCTG GTGAAAAGAA AAACCCACCT	6000
GGCGCCCAAT ACGCAAACCG CCTCTCCCCG CGCGTTGGCC GATTCACTAA TGAGCTGGC	6060
ACGACAGGTT TCCCCACTGG AAAGCGGGCA GTGAGCGCAA CGCAATTAAAT GTGAGTTAGC	6120
TCACTCATTA GGCACCCAG GCTTTACACT TTATGCTTCC GGCTCGTATG TTGTGTGGAA	6180
TTGTGAGCGG ATAACAATTT CACACGGCTC ACTTGGCACT GGCGTGGTT TTACAACGTC	6240
GTGACTGGGA AAACCCCTGGC GTTACCCAAG CTTTGTACAT GGAGAAAATA AAGTGAAACA	6300
AAGCACTATT GCACTGGCAC TCTTACCGTT ACCGTTACTG TTTACCCCTG TGACAAAAGC	6360
GGCCCAGGTC CAGCTGCTGG AGTCAGGCCT ATTGTGCCCA GGGGATTGTA CTAGTGGATC	6420
CTAGGCTGAA GGCGATGACC CTGCTAAGGC TGCATTCAAT AGTTTACAGG CAAAGTGGCTAC	6480
TGAGTACATT GGCTACGCTT GGGCTATGGT AGTAGTTATA GTTGGTGCTA CCATAGGGAT	6540

TAAATTATTC AAAAAGTTTA CGAGCAAGGC TTGTTAACGA ATAGCGAAGA GGCCCCGACCC	6600
GATGGCCCTT CCCAACAGTT GCGCAGCCTG AATGGCGAAT GGGCCTTGCG CTGGTTTCCG	6660
GCACCCAGAAG CGGTGCCGGA AAGCTGGCTG GAGTGGATC TTGCTGAGGC CGATAACGGTC	6720
GTCGTCCCTT CAAACTGGCA GATGCACGGT TAGCATGCC CCATCTACAC CAACGTAACC	6780
TATCCCATTAA CGGTCAATCC GCCGTTTGTT CCCACGGAGA ATCCGACGGG TTGTTACTCG	6840
CTCACATTTA ATGTTGATGA AAGCTGGCTA CAGGAAGGCC AGACGCGAAT TATTTTTGAT	6900
GGCGTTCCCTA TTGGTTAAAA AATGAGCTGA TTTAACAAAA ATTTAACCGG AATTTTAACA	6960
AAATATTAAC GTTTACAATT TAAATATTTG CTTATACAAT CTTCTGTGTT TTGGGGCTTT	7020
TCTGATTATC AACCGGGGTA CATATGATTG ACATGCTAGT TTTACGATTA CCGTTCATCG	7080
ATTCTCTTGT TTGCTCCAGA CTCTCAGGCA ATGACCTGAT AGCCCTTGTA GATCTCTGAA	7140
AAATAGCTAC CCTCTCCGGC ATTAATTTAT CAGCTAGAAC GGTTGAATAT CATATTGATG	7200
GTGATTTGAC TGTCTCCGGC CTTTCTCACC CTTTTGAATC TTTACGTTACA CATTACTCAG	7260
GCATTGCATT TAAAATATAT GAGGGTTCTA AAAATTTTA TCCTTGGGTT GAAATAAAGG	7320
CTTCTCCGGC AAAAGTATTAA GAGGGTCATA ATGTTTTGG TAGAACGGAT TTAGCTTTAT	7380
GCTCTGAGGC TTATGCTT AATTTTGCTA ATTCTTGCC TTGCTGTAT GATTTATTGG	7440
ACGTT	7445

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 7317 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: both
 - (D) TOPOLOGY: circular

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

AATGCTACTA CTATTAGTAG AATTGATGCC ACCTTTTCAG CTCGGGCCCC AAATGAAAAT	60
ATAGCTAAAC AGGTTATTGA CCATTTGCGA AATGTATCTA ATGGTCAAAC TAAATCTACT	120
CGTTCCAGA ATTGGGAATC AACTGTTACA TGGAATGAAA CTTCCAGAGA CCGTACTTTA	180
TTGGCATATT TAAAACATGT TGAGCTACAG CACCAAGATTG AGCAATTAAG CTCTAAGCCA	240
TCCGAAAAAA TGACCTCTTA TCAAAAGGAG CAATTAAAGG TACTCTCTAA TCCIGACCTG	300
TTGGAGTTTG CTTCCGGTCT GGTTCGCTTT GAAGCTCGAA TAAACAGCGG ATATTTGAAG	360
TCTTTCGGGC TTCCTCTTAA TCTTTTGAT GCAATCGCT TTGCTCTGA CTATAATAGT	420
CAGGGTAAAG ACCTGATTT TGATTTATGG TCATTCTCGT TTTCTGAACG GTTTAAAGCA	480
TTTGAGGGGG ATTCAATGAA TATTTATGAC GATTCCGCAG TATTGGACCC TATCCAGTGT	540
AAACATTTTA CTATTACCCC CTCTGGCAAA ACTTCTTTG CAAAAGCCTC TCGCTATTT	600
GGTTTTTATC GTCGTCTGGT AAACGAGGGT TATGATAGTG TTGCTCTTAC TATGCTCGT	660
AATTCCCTTTT GGCGTTATGT ATCTGCACTTA GTTGAATGTG GTATTCTAA ATCTCAACTG	720

ATGAATCTTT	CTACCTGTAA	TAATGTTGTT	CCGTTAGTTC	GTTTTATTAA	CGTAGATT	780
TCTTCCCAAC	GTCCTGACTG	GTATAATGAG	CCAGTTCTTA	AAATCGCATA	AGGTAATTCA	840
CAATGATTAA	AGTTGAAATT	AAACCATCTC	AAGCCCAATT	TACTACTCGT	TCTGGTGT	900
CTCGTCAGGG	CAAGCCTAT	TCACTGAATG	AGCAGCTTG	TTACGTTGAT	TTGGGTAATG	960
AATATCCGGT	TCTTGTCAAG	ATTACTCTTG	ATGAAGGTCA	GCCAGCCTAT	GCGCCTGGTC	1020
TGTACACCGT	TCATCTGTCC	TCTTCAAAG	TTGGTCAGTT	CGGTTCCGTT	ATGATTGACC	1080
GTCTGCGCCT	CGTTCCGGCT	AACTAACATG	GAGCAGGTG	CGGATTCGA	CACAATTAT	1140
CAGGCATG	TACAAATCTC	CGTTGTACTT	TGTTTCGCGC	TTGGTATAAT	CGCTGGGGT	1200
CAAAGATGAG	TGTTTAGTG	TATTCTTTCG	CCCTCTTCGT	TTTAGGTTGG	TGCCCTCGTA	1260
GTGGCATTAC	GTATTTACC	CGTTTAATGG	AAACTCCTC	ATGAAAAAGT	CTTTAGTCCT	1320
CAAAGCCTCT	GTAGCCGTTG	CTACCCCTCGT	TCCGATGCTG	TCTTCGCTG	CTGAGGGTGA	1380
CGATCCCGA	AAAGCCGCT	TTAACTCCGT	GCAAGCCTCA	CGCACCGAAT	ATATCGGTTA	1440
TGCGTGGCG	ATGGTTGTTG	TCATTGTCGG	CGGAACATAC	GGTATCAAGC	TGTTTAAGAA	1500
ATTCACTCG	AAAGCAAGCT	GATAAACCGA	TACAATTAAA	GGCTCCTTTT	GGAGCCTTTT	1560
TTTTGGAGA	TTTCAACGT	GAAAAAATTA	TTATTGCAA	TTCCCTTAGT	TGTTGCTTT	1620
TATTCTCACT	CCGCTGAAAC	TGTTGAAAGT	TGTTAGCAA	AACCCATAC	AGAAAATTCA	1680
TTTACTAACG	TCTGGAAAGA	CGACAAAAC	TTAGATCGTT	ACGCTAACTA	TGAGGGTTGT	1740
CTGTGGAATG	CTACAGGGGT	TGTAGTTGT	ACTGGTGACG	AAACTCAGTG	TTACGGTACA	1800
TGGGTTCCCTA	TTGGGTTG	TATCCCTGAA	AATGAGGGTG	GTGGCTCTGA	GGGTGGCGGT	1860
TCTGAGGGTG	GCGGTTCTGA	GGGTGGCGGT	ACTAAACCTC	CTGAGTACCG	TGATACACCT	1920
ATTCCGGGCT	ATACTTATAT	CAACCCCTCG	GACGGCAGT	ATCCGCTGG	TACTGAGCAA	1980
AACCCCCGCTA	ATCCTAATCC	TTCTCTTGAG	GAGTCTCAGC	CTCTTAATAC	TTTCATGTT	2040
CAGAATAATA	GGTTCCGAAA	TAGGCAGGGG	GCATTAAC	TTTATACGGG	CACTGTTACT	2100
CAAGGGCACTG	ACCCCGTTAA	AACTTATTAC	CACTACACTC	CTGTATCATC	AAAAGCCATG	2160
TATGACGCTT	ACTGGAACGG	AAAATTCAA	GAATGCGCTT	TCCATTCTGG	CTTTAATGAA	2220
GATCCATTCTG	TTTGTGAATA	TCAAGGCCAA	TGCTCTGACC	TGCCTCAACC	TCCTGTCAAT	2280
GCTGGCGCG	GCTCTGGTGG	TGGTTCTGGT	GGCGGCTCTG	AGGGTGGTGG	CTCTGAGGGT	2340
GGCGGTTCTG	AGGGTGGCGG	CTCTGACGGG	GGCGGTTCCG	GTGGTGGCTC	TGGTTCCGGT	2400
GAITTTGATT	ATGAAAAGAT	GGCAAACGCT	AATAAGGGGG	CTATGACCGA	AAATGCCGAT	2460
GAAAACGCGC	TACAGTCTGA	CGCTAAAGGC	AAACTTGATT	CTGTCGCTAC	TGATTACGGT	2520
GCTGCTATCG	ATGGTTTCAT	TGGTGACGTT	TCCGGCTTG	CTAATGGTAA	TGGTGCTACT	2580
GGTGATTTTG	CTGGCTCTAA	TTCCCAAATG	GCTCAAGTGG	GTGACGGTGA	TAATTCACT	2640
TTAATGAATA	ATTTCCGTCA	ATATTTACCT	TCCCTCCCTC	AATCGGTGA	ATGTCGCCCT	2700
TTTGTCTTAA	GCGCTGGTAA	ACCATATGAA	TTTCTATTG	ATTGTGACAA	AATAAACTTA	2760

TTCCGTGGTG TCTTTCGTT TCTTTATAT GTGCCACCT TTATGTATGT ATTTCTACG	2820
TTTGCTAACAA TACTCGTAA TAAGGAGTCT TAATCATGCC AGTTCTTTG GGTATTCCGT	2880
TATTATTGGG TTTCCTCGGT TTCCCTCTGG TAACCTTGTG CCGCTATCTG CTTACTTTG	2940
TTAAAAAGGG CTTCGTAAG ATAGCTATTG CTATTCGATT GTTCTTGCT CTTATTATTG	3000
GGCTTAACTC AATTCTGTG GGTTATCTCT CTGATATTAG CGCTCAATTA CCCTCTGACT	3060
TTGTTCAAGGG TGTTCAAGTTA ATTCTCCGT CTAATGGCT TCCCTGTTT TATGTTATTG	3120
TCTCTGAAA CGCTGCTATT TTCATTTTG ACGTTAAACA AAAAATCGTT TCTTATTTGG	3180
ATTGGGATAAA ATAATATGGC TGTTTATTTT GTAACTGGCA AATTAGGCTC TGGAAAGACG	3240
CTCGTTAGCG TTGGTAAGAT TCAGGATAAA ATTGTACCTG GGTGCAAAT AGCAACTAAT	3300
CTTGATTTAA GGCTTAAAAA CCTCCCGCAA GTCGGGAGGT TCGCTAAAC GCCTCGCGTT	3360
CTTAGAATAC CGGATAAGCC TTCTATATCT GATTTGCTTG CTATTGGCG CGGTAATGAT	3420
TCCTACGATG AAAATAAAA CGGCTTGCCT GTTCTCGATG AGTCCGGTAC TTGGTTAAAT	3480
ACCCGTTCTT GGAATGATAA GGAAAGACAG CCGATTATTG ATTGGTTCT ACATGCTCGT	3540
AAATTAGGAT GGGATATTAT TTTCCTGTGTT CAGGACTTAT CTATTGTTGA TAAACAGGCG	3600
CGTTCTGGAT TAGCTGAACA TGTTGTTTAT TGTCGTCGTC TGGACAGAAT TACTTTACCT	3660
TTTGTGGTA CTTTATATTC TCTTATTACT GGCTCGAAAA TGCCTCTGCC TAAATTACAT	3720
TTTGGCGTIG TAAATATGG CGATTCTCAA TTAAGCCTA CTGTTGAGCG TTGGTTTAT	3780
ACTGGTAAGA ATTGTATAA CGCATATGAT ACTAAACAGG CTTTTCTAG TAATTATGAT	3840
TCCGGTGTGTT ATTCTTATTT AACGGCTTAT TTATCACACG GTCGGTATTT CAAACCATT	3900
AATTTAGGTC AGAACATGAA GCTTACTAAA ATATATTGAA AAAAGTTTC ACGCGTTCTT	3960
TGTCTTGCAGA TTGGATTGTC ATCAGCATT ACATATAGTT ATATAACCCA ACCTAACCG	4020
GAGGTTAAAA AGCTACTCTC TCAGACCTAT GATTTGATA AATTCACTAT TGACTCTCT	4080
CAGCGTCTTA ATCTAAGCTA TCGCTATGTT TTCAAGGGATT CTAAGGGAAA ATTAATTAAT	4140
AGCGACGATT TACAGAAGCA AGGTTATTCA CTCACATATA TTGATTTATG TACTGTTCC	4200
ATTAAAAAAG GTAATTCAAA TGAAATTGTT AAATGTAATT AATTGTTTT TCTTGATGTT	4260
TGTTTCATCA TCTTCCTTTG CTCAGGTAAT TGAAATGAAT AATTGCGCTC TGGCGGATT	4320
TGTAACCTGG TATTCAAAGC AATCAGGCGA ATCCGTTATT GTTCTCCCG ATCTAAAAGG	4380
TACTGTTACT GTATATTCTAT CTGACGTTAA ACCTGAAAAT CTAGCAATT TCTTATTT	4440
TGTTTACGT GCTAATAATT TTGATATGGT TGGTTCAATT CCTTCATAA TTCAAGAGTA	4500
TAATCCAAAC AATCAGGATT ATATTGATGA ATTGCCATCA TCTGATAATC AGGAATATG	4560
TGATAATTCC GCTCCTCTG GTGGTTCTT TGTTCCGAA AATGATAATG TTACTCAAAC	4620
TTTTAAAATT AATAACGTTG GGGCAAAGGA TTAAATACGA GTTGTGGAAT TGTTTGTAAA	4680
GTCTAATACT TCTAAATCCT GAAATGTATT ATCTATTGAC GGCTCTAATC TATTAGTTGT	4740
TAGTGCACCT AAAGATATTI TAGATAACCT TCCCTCAATTG CTTTCTACTG TTGATTTGCC	4800

AACTGACCGAG ATATTGATTG AGGGTTTGAT ATTTGAGGTT CAGCAAGGTG ATGCTTTAGA	4860
TTTTCATTT GCTGCTGGCT CTCAGCGTGG CACTGTTGCA GGCGGTGTTA ATACTGACCG	4920
CCTCACCTCT GTTTTATCTT CTGCTGGTGG TTGCTTCGGT ATTTTTAATG GCGATGTTTT	4980
AGGGCTATCA GTTCGGCGAT TAAAGACTAA TAGCCATTCA AAAATATTGT CTGTGCCACG	5040
TATTCTTACG CTTTCAGGTC AGAAGGGTTC TATCTCTGTT GGCCAGAATG TCCCTTTAT	5100
TACTGGTGGT GTGACTGGTG AATCTGCCAA TGTAAATAAT CCATTTCAGA CGATTGAGCG	5160
TCAAAATGTA GGTATTCCA TGAGCGTTT TCCTGTTGCA ATGGCTGGCG GTAATATTGT	5220
TCTGGATATT ACCAGCAAGG CCGATAGTTT GAGTTCTTCT ACTCAGGCAA GTGATGTTAT	5280
TACTAATCAA AGAAGTATTG CTACAACGGT TAATTTGCGT GATGGACAGA CTCTTTACT	5340
CGGTGGCCTC ACTGATTATA AAAACACTTC TCAAGATTCT GGCGTACCGT TCCTGTCTAA	5400
AATCCCTTTA ATCGGCCTCC TGTAGCTC CCGCTCTGAT TCCAACGAGG AAAGCACGTT	5460
ATACGTGCTC GTCAAAGCAA CCATAGTACG CGCCCTGTAG CGGGCGATTAA AGCGCGGGCG	5520
GTGTGGTGGT TACGGCGAGC GTGACCGCTA CACTTGGCAG CGCCCTAGCG CCCGCTCCCT	5580
TCGCTTTCTT CCCCTTCCCTT CTCGCCACGT TCGCCGGCTT TCCCCGTCAA GCTCTAAATC	5640
GGGGGCTCCC TTAGGGTTC CGATTTAGTG CTTTACGGCA CCTCGACCCCC AAAAAACTTG	5700
ATTTGGGTGA TGGTTCACGT AGTGGGCCAT CGCCCTGATA GACGGTTTTT CGCCCTTGA	5760
CGTTGGAGTC CACGTTCTTT AATAGTGGAC TCTTGTCCA AACTGGAACA ACACCTAAC	5820
CTATCTGGG CTATTCTTT GATTATAAG GGATTTGCC GATTTCGGAA CCACCATCAA	5880
ACAGGATTTT CGCCTGCTGG GGCAAAACCAG CGTGGACCCG TTGCTGCAAC TCTCTCAGGG	5940
CCAGGGCGTG AAGGGCAATC AGCTGTTGCC CGTCTCGCTG GTGAAAAGAA AAACCACCC	6000
GGCGCCCAAT ACGCAAACCG CCTCTCCCCG CGCGTTGGCC GATTCAATTAA TGCAGCTGGC	6060
ACGACAGGTT TCCCGACTGG AAAGCGGGCA GTGAGGCCAA CGCAATTAAAT GTGAGTTAGC	6120
TCACTCATTA GCCACCCCCAG CCTTTACACT TTATGCTTCC GGCTCGTATG TTGTGTGGAA	6180
TTGTGAGCGG ATAACAATT TACACGCCAA GGAGACAGTC ATAATGAAAT ACCTATTGCC	6240
TACGGCAGCC GCTGGATTGT TATTACTCGC TGCCCAACCA GCCATGGCCG AGCTCGTGAT	6300
GACCCAGACT CCAGATATCC AACAGGAATG AGTGTAAATT CTAGAACCGC TCACTTGGCA	6360
CTGGCCGTGG TTTACAACG TCGTGAATGG GAAAACCCCTG CGGTACCCA AGCTTAATCG	6420
CCTTGCAGAA TTCCCTTTCG CGAGCTGGCG TAATAGGCCAA GAGGCCCGCA CGATCGCCC	6480
TTCCCAACAG TTGGCAGCC TGAATGGCGA ATGGCGCTTT GCCTGGTTTC CGGCACCAAGA	6540
AGCGGTGCCG GAAAGCTGGC TGGAGTGCAGA TCTTCCGTAG GCCGATACGG TCGTCGTCCC	6600
CTCAAACCTGG CAGATGCCAGC GTTACCGATGC GCCCATCTAC ACCAACGTAA CCTATCCC	6660
TACGGTCAAT CGGCGCTTTC TTCCCCACGGA GAATCCGACG GGTGTTACT CGCTCACATT	6720
TAATGTTGAT GAAAGCTGGC TACAGGAAGG CCAGACGCCA ATTATTTTG ATGGCGTTCC	6780
TATTGGTTAA AAAATGAGCT GATTAAACAA AAATTTAACG CGAATTTAA CAAAATATTA	6840

ACGTTTACAA TTTAAATATT TGCTTATACA ATCTTCCTGT TTTGGGGCT TTCTGATTA	6900
TCAACCGGGG TACATATGAT TGACATGCTA GTTTACGGAT TACCGTCAT CGATTCTCTT	6960
GTTTGCTCCA GACTCTCAGG CAATGACCTG ATAGCCTTG TAGATCTCTC AAAAATAGCT	7020
ACCCCTCTCCG GCATTAATTT ATCAGCTAGA ACGGTTGAAT ATCATATTGA TGGTGATTG	7080
ACTGTCTCCG GCCTTGTCA CCCTTTGAA TCTTACCTA CACATTACTC AGGCATTGCA	7140
TTTAAAATAT ATGAGGGTTC TAAAAAATTAT TATCCTGCG TTGAAATAAA GGCTTCTCCC	7200
GCAAAAGTAT TACAGGGTCA TAATGTTTTT GGTACAACCG ATTTAGCTTT ATGCTCTGAG	7260
GCTTTATTGC TTAATTTCG TAATTCTTG CCTTGCCTGT ATGATTATT GGATGTT	7317

(2) INFORMATION FOR SEQ ID NO:3:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 7729 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: both
 - (D) TOPOLOGY: circular

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

AATGCTACTA CTATTAGTAG AATTGATGCC ACCTTTCAAG CTGGCGCCCC AAATGAAAAT	60
ATAGCTAAC AGGTTATTGA CCATTTGCGA AATGTATCTA ATGGTCAAAC TAAATCTACT	120
CGTTCGAGA ATTGGGAATC AACTGTTACA TGGAATGAAA CTTCCAGACA CCGTACTTTA	180
GTTGCATATT TAAAACATGT TGAGCTACAG CACCAGATTG AGCAATTAAG CTCTAAGCCA	240
TCTGAAAAAA TGACCTCTTA TCAAAAGGAG CAATTAAAGG TACTCTCTAA TCCTGACCTG	300
TTGGAGTTTG CTTCCGGTCT GGTTCGCTTT GAAGCTCGAA TTAAAACCGG ATATTGAAAG	360
TCTTTCGGGC TTCCCTTAA TCTTTTGAT GCAATCCGCT TTGCTCTGA CTATAATAGT	420
CAGGGTAAAG ACCTGATTTT TGATTTATGG TCATTCTCGT TTCTGAACG GTTTAAAGCA	480
TTTGAGGGG ATTCAATGAA TATTTATGAC GATTCCGCAG TATTGGACGC TATCCAGTCT	540
AAACATTTTA CTATTACCCC CTCTGGCAAA ACTTCTTTG CAAAAGCCTC TCGCTATT	600
GGTTTTTATC GTCGTCTGGT AAACGAGGGT TATGATACTG TTGCTCTTAC TATGCCCTCGT	660
AATTCCCTTT GGCGTTATGT ATCTGCATTA GTTGAATGTG GTATTCTAA ATCTCAACTG	720
ATGAATCTTT CTACCTGTAA TAATGTTGTT CCGTTAGTTG GTTTTATTAA CGTAGATT	780
TCTTCCCAAC GTCCGTACTG GTATAATGAG CCAGTTCTTA AAATGGCATA AGGTAATTCA	840
CAATGATTAA AGTTGAAATT AAACCATCTC AAGCCCAATT TACTACTCGT TCTGGTGT	900
CTCGTCAGGG CAAGCCCTAT TCACGTGAATG AGCAGCTTG TTACGTTGAT TTGGGTAATG	960
AATATCCGGT TCTTGTCAAG ATTACTCTTG ATGAAGGTCA GCCAGCCTAT GCGCCTGGTC	1020
TGTACACCGT TCATCTGTCC TCTTCAAAAG TTGGTCAGTT CGGTTCCCTT ATGATTGACC	1080
GTCTGGCCT CGTTCCGGCT AAGTAACATG GAGCAGGTGG CGGATTTGCA CACAATTAT	1140
CAGGGCATGA TACAAATCTC CGTTGTACTT TGTTTCCGCC TTGGTATAAT CGCTGGGGT	1200

CAAAGATGAG TGTGTTAGTG TATTCTTCG CCTCTTCGT TTTAGGTGG TGCCCTCGTA	1260
GTGGCATTAC GTATTTACC CGTTAATGG AAACCTCCTC ATGAAAAAGT CTTTAGTCCT	1320
CAAAGCCTCT GTAGCCGTTG CTACCCCTCGT TCC..TGCTG TCTTCGCTG CTGAGGGTGA	1380
GGATCCCAGCA AAAGCGGCCT TTAACTCCCT GCAAGCCTCA GCGACCGAAT ATATCGGTAA	1440
TGCGTGGGGCG ATGGTTGTTG TCATTGTCGG CGCAACTATC GGTATCAAGC TGTTAACGAA	1500
ATTCACCTCG AAAGCAAGCT GATAAACCGA TACAATTAAA GGCTCCTTT GGAGCCTTT	1560
TTTTGGAGA TTTTCAACGT GAAAAAATTA TTATTGGAA TTCCTTAGT TGTTCCCTTC	1620
TATTCTCACT CCGCTGAAAC TGTGAAAGT TGTTAGGAA AACCCCATAC AGAAAATTCA	1680
TTTACTAACG TCTGGAAAGA CGACAAAATC TTAGATCGTT ACCCTAACTA TGAGGGTTGT	1740
CTGTGGAATG CTACAGGGCGT TGTAGTTGT ACTGGTGACG AAACTCAGTG TTACGGTACA	1800
TGGGTTCCCTA TTGGGCTTGC TATCCCTGAA AATGAGGGTG GTGGCTCTGA GGGTGGCGGT	1860
TCTGAGGGTG GCGGTTCTGA GGGTGGCGGT ACTAAACCTC CTGAGTACGG TGATACACCT	1920
ATTCCGGGCT ATACTTATAT CAACCCCTCTC GACGGCACTT ATCCGCTGG TACTGAGCAA	1980
AACCCCGCTA ATCCTAATCC TTCTCTTGAG GAGTCTCAGC CTCTTAATAC TTTCATGTTT	2040
CAGAATAATA GGTTCCGAAA TAGGCAGGGGG GCATTAACGT TTATACGGG CACTGTTACT	2100
CAAGGCACGT ACCCCGTTAA AACTTATTAC CAGTACACTC CTGTATCATC AAAAGCCATG	2160
TATGACGCTT ACTGGAACGG TAAATTGAGA GACTGCGCTT TCCATTCTGG TTATTAATGAA	2220
GATCCATTCCG TTGTGAATA TCAAGGCCAA TCGTCTGACC TGCCTGAACC TCCTGTCAAT	2280
GCTGGCGGGCG GCTCTGGTGG TGTTCTGGT GGCGGCTCTG AGGGTGGTGG CTCTGAGGGT	2340
GGCGGTTCTG AGGGTGGCGG CTCTGAGGGA GGCGGTTCCG GTGGTGGCTC TGTTCCGGT	2400
GATTTGATT ATGAAAAGAT GGCAAAACGCT AATAAGGGGG CTATGACCGA AAATGCCGAT	2460
AAAAACGGCG TACAGTCTGA CGCTAAAGGC AAACCTGATT CTGTCGCTAC TGATTACGGT	2520
GCTGCTATCG ATGGTTTCAT TGTTGACGTT TCCGGCTTG CTAATGGTAA TGTTGCTACT	2580
GGTGATTTG CTGGCTCTAA TTCCCAAATG GCTCAAGTCG GTGACGGTGA TAATTCCACCT	2640
TTAATGAATA ATTTCCGTCA ATATTTACCT TCCCTCCCTC AATCGGTGA ATGTCGCCCT	2700
TTTGTCTTAA GCGCTGGTAA ACCATATGAA TTTTCTATTG ATTGTGACAA AATAAAACTTA	2760
TTCCGTGGTG TCTTTCGGT TCTTTTATAT GTTGCACCT TTATGTATGT ATTTTCTACG	2820
TTTGCTAACCA TACTGGTAA TAAGGAGTCT TAATCATGCC AGTTCTTTG GGTATTCCGT	2880
TATTATTGCG TTCCCTCGGT TTCCCTCTGG TAACCTTGTT CGGCTATCTG CTTACTTTTC	2940
TTAAAAAGGG CTTCGTAAG ATAGCTATTG CTATTCATT GTTTCTGCT CTTATTATTG	3000
GGCTTAACTC AATTCTGTG GGTTATCTCT CTGATATTAG CGCTCAATTA CCCTCTGACT	3060
TTGTTCAAGGG TGTTCACTTA ATTCTCCCGT CTAATGGCT TCCCTGTTTT TATGTTATTG	3120
TCTCTGTAAA GGCTGCTATT TTCAATTG ACGTTAAACA AAAAATCGTT TCTTATTG	3180
ATTGGGATAAA ATAATATGGC TGTTTATTIT GTAACTGGCA AATTAGGCTC TGGAAAGACG	3240

CTCGTTAGCG TTGGTAAGAT TCAGGATAAA ATTGTAGCTG GGTGAAAAT AGCAACTAAT	3300
CTTGATTTAA GGCTCAAAA CCTCCCGCAA GTCGGGAGGT TCGCTAAAC GCCTCGCGTT	3360
CTTAGAATAAC CGGATAAGCC TTCTATATCT GATTGCTTGC CTATTGGCG CGGTAATGAT	3420
TCCTACGATG AAAATAAAAA CGGCTTGCTT GTTCTCGATG AGTGCCTGAC TTGGTTTAAT	3480
ACCCGTTCTT GGAATGATAA GGAAAGACAG CCGATTATTG ATGGGTTCT ACATGCTCGT	3540
AAATTAGGAT GGGATATTAT TTTCTTGTT CAGGACTTAT CTATTGTTGA TAAACAGGCG	3600
CGTTCTGCAT TAGCTGAACA TGTGTTTAT TGTCGTCGTC TGGACAGAAT TACTTTACCT	3660
TTTGTGGTA CTTTATATTC TCTTATTACT GGCTCGAAA TGCCTCTGCC TAAATTACAT	3720
GTTGGCGTTG TAAATATGG CGATTCTGAA TTAAGCCTA CTGTTGAGCG TTGGCTTTAT	3780
ACTGGTAAGA ATTTGTATAA CGCATATGAT ACTAACACGG CTTTTCTAG TAATTATGAT	3840
TCCGGTGTIT ATTCTTATTT AACGCCTTAT TTATCACACG GTCGGTATTT CAAACCATT	3900
AATTTAGGTC AGAAGATGAA GCTTACTAAA ATATATTGA AAAAGTTTC ACGCGTTCTT	3960
TGTCTTGCAG TTGGATTTGC ATCAGGATT ACATATACTT ATATAACCCA ACCTAACCG	4020
CAGGTTAAA AGGTAGTCTC TCAGACCTAT GATTTGATA AATTCACTAT TGAACCTCT	4080
CAGCGTCTTA ATCTAAGCTA TCGCTATGTT TTCAAGGATT CTAAGGGAAA ATTAATTAAT	4140
ACCGACGATT TACAGAAGCA AGGTTATTCA CTCACATATA TTGATTTATG TACTGTTCC	4200
ATTAaaaaaaAG GTAATTCAA TGAATTGTT AAATGTAATT AATTGTTTCT TCTTGATGTT	4260
TGTTTCATCA TCTTCTTTG CTCAGGTAAT TGAAATGAAT AATTGGCTC TGCGCGATT	4320
TGTAACCTGG TATTCAAAGC AATCAGGCGA ATCCGTTATT GTTCTCCCCG ATGTAAGG	4380
TACTGTTACT GTATATTCACT CTGACGTTAA ACCTGAAAAT CTACGCAATT TCTTTATTTC	4440
TGTTTTACGT GCTAATAATT TTGATATGGT TGGTTCAATT CCTTCCATAA TTCAGAAGTA	4500
TAATCCAAAC AATCAGGATT ATATTGATGA ATTGCCATCA TCTGATAATC AGGAATATGA	4560
TGATAATTCC GCTCCTCTG GTGGTTCTT TGTTCCCAA AATGATAATG TTACTCAAAC	4620
TTTTAAAATT AATAACGTTG GGGCAAAGGA TTTAATACGA GTTGTGAAAT TGTTGTAAA	4680
GTCTAATACT TCTAAATCCT CAAATGTATT ATCTATTGAC GGCTCTAAC TATTAGTTGT	4740
TAGTGCACCT AAAGATATT TAGATAACCT TCCCTCAATT CTTTCTACTG TTGATTTGCC	4800
AACTGACCAAG ATATTGATTG AGGGTTGAT ATTTGAGGTT CAGCAAGGTG ATGCTTTAGA	4860
TTTTTCATTG GCTGCTGGCT CTCAGCGTGG CACTGTTGCA GGCGGTGTTA ATACTGACCG	4920
CCTCACCTCT GTTTATCTT CTGCTGGTGG TTGTTGGT ATTGTTAATG GCGATGTTT	4980
AGGGCTATCA GTTCCGGCAT TAAAGACTAA TAGCCATTCA AAAATATTGT CTGTGCCACG	5040
TATTCTTACG CTTTGAGGTC AGAAGGGTTC TATCTCTGTT GGCCAGAATG TCCCTTTAT	5100
TACTGGTGGT GTGACTGGTG AATCTGCCAA TGAAATAAT CCATTCAGA CGATTGAGCG	5160
TCAAAATGTA GGTATTCCA TGAGCGTTT TCCTGTTGCA ATGGCTGGCG GTAATATTGT	5220
TCTGGATATT ACCAGCAAGG CCGATAGTTT GAGTTCTCT ACTCAGGCAA GTGATGTTAT	5280

TACTAATCAA AGAAAGTATTG CTACAACGGT TAATTTGCGT GATGGACAGA CTCTTTACT	5340
CGGTGGCCTC ACTGATTATA AAAACACTTC TCAAGATTCT GGCGTACCGT TCCTGTCTAA	5400
AATCCCTTTA ATCGGCCCTCC TGTTTAGCTC CCGCTCTGAT TCCAACGAGG AAAGCACGTT	5460
ATACGTGCTC GTCAAAGCAA CCATACTAGG CGCCCTGTAG CGGGCGCATTA AGCGCGGGGG	5520
GTGTGGTGGT TACGGGCAGC GTGACCGCTA CACTTGCCAG CGCCCTAGCG CCCGCTCCTT	5580
TCGCTTTCTT CCCCTCTTT CTGGCCACGT TCGCCGGCTT TCCCCGTCAA GCTCTAAATC	5640
GGGGGCTCCC TTTAGGGTTC CGATTTAGTC CTTTACGGCA CCTCGACCCC AAAAAACTTG	5700
ATTTGGGTGA TGGTTCACGT AGTGGGCCAT CGCCCTGATA GACGGTTTT CGCCCTTTGA	5760
CGTTGGAGTC CACGTCTTT AATAGTGGAC TCTTGTCTCA AACTGGAACA ACACCTAAC	5820
CTATCTCGGG CTATTCTTT GATTIATAAG GGATTTGCC GATTTCGGAA CCACCATCAA	5880
ACAGGATTTT CGCCTGCTGG GGGAAACCCAG CGTGGACCCGC TTGCTGCAAC TCTCTCAGGG	5940
CCAGGGGGTG AAGGGCAATC AGCTGTTGCC CGTCTCGCTG GTGAAAAGAA AAACCCACCT	6000
GGCGCCCAAT ACGGAAACCG CCTCTCCCCG CGCGTTGGCC GATTCACTAA TGGAGCTGGC	6060
ACGACAGGTT TCCCAGCTGG AAAGCGGGCA GTGAGGCCAA CGCAATTAAT GTGAGTTAGC	6120
TCACTCACTA GGCACCCAG GCTTTACACT TTATGCTTCC GGCTCGTATG TTGTGTGGAA	6180
TTGTGAGCGG ATAACAATTT CACACGGCTC ACTTGGCACT GGCGTCGTT TTACAACGTC	6240
GTGACTGGGA AAACCCCTGGC GTTACCCAAG CTTTGTACAT GGAGAAAATA AAGTGAACACA	6300
AAGCACTATT GCACCTGGCAC TCTTACCGTT ACTGTTTACC CCTGTGGCAA AAGCCCAGGT	6360
CCAGCTGCTC GAGTCGGTCT TCCCCCTGGC ACCCTCTCC AAGAGCACCT CTGGGGGCAC	6420
AGCGGGCCCTG GGCTGGCTGG TCAAGACTAA TTCCCCGAAC CGGTGACGGT GTCGTGGAAC	6480
TCAGGGGCCCG TGACCAGCGG CGTGCACACC TTCCCCGGCTG TCCTACAGTC CTCAGGACTC	6540
TACTCCCTCA CGAGCGTGGT GACCGTGGCC TCCAGCACCT TGGCACCCCA GACCTACATC	6600
TGCAACGTGA ATCACAAGCC CAGAACACCC AAGGTGGACA AGAAAGCAGA GCCCAAATCT	6660
TGTACTAGTG GATCCTACCC GTACGACGTT CC ACTAGG CTTCTTAGGC TGAAGGGCGAT	6720
GACCCCTGCTA AGGCTGCATT CAATAGTTA CAGGCAAGTG CTACTGAGTA CATTGGCTAC	6780
GCTTGGGCTA TGGTAGTAGT TATAGTTGGT GCTACCATAG GGATTAATT ATTCAAAAG	6840
TTTACGAGCA AGGCTTCITA AGCAATAGCG AAGAGGCCCG CACCGATCCG CCTTCCCAAC	6900
AGTTGCGCAG CCTGAATGGC GAATGGCGCT TTGCGTGGTT TCCGGCACCA GAAGCGGTGC	6960
CGGAAAGCTG GCTGGAGTGC GATCTTCCTG AGGCGATAC GGTGTCGTC CCCTCAAAC	7020
GGCAGATGCA CGGTTACGAT GCGCCGATCT ACACCAACGT AACCTATCCC ATTACGGTCA	7080
ATCCGCCGTT TGTTCACCGG GAGAATCCGA CGGGTTGTTA CTGCTCACA TTTAATGTTG	7140
ATGAAAGCTG GCTACAGGAA GCCCAGACGC GAATTATTT TGATGGCGTT CCTATTGGTT	7200
AAAAAAATGAG CTGATTTAAC AAAAATTAA CGCGAATTTT AACAAAATAT TAACGTTTAC	7260
AATTAAATA TTTGCTTATA CAATCTTCCT GTTTTGGGG CTTTTCTGAT TATCAACCGG	7320

GGTACATATG ATTGACATGC TAGTTTACG ATTACCGITC ATCGATTCTC TTGTTTGCTC	7380
CAGACTCTCA GGCAATGACC TGATAGCCTT TGTAGATCTC TCAAAAATAG CTACCCCTCTC	7440
CGGCATTAAT TTATCAGCTA GAACGGTTGA ATATCATATT GATGGTGATT TGACTGTCTC	7500
CGGCCCTTTCT CACCCCTTTG AATCTTTACC TACAGATTAC TCAGGCATTG CATTAAAT	7560
ATATGAGGGT TCTAAAAATT TTTATCCTTG CGTTGAAATA AAGGCTTCTC CCGCAAAAGT	7620
ATTACAGGGT CATAATGTTT TTGGTACAAC CGATTTAGCT TTATGCTCTG AGGCTTTATT	7680
GCTTAATTTT GCTAATTCTT TGCCCTGGCT GTATGATTAA TTGGACGTT	7729

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 7557 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: both
 - (D) TOPOLOGY: circular

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

AATGCTACTA CTATTAGTAG AATTGATGCC ACCTTTICAG CTCGGCCCC AAATGAAAAT	60
ATAGCTAAC AGGTTATTGA CCATTGCGA AATGTATCTA ATGGTCAAAC TAAATCTACT	120
CGTTCGCAGA ATTGGGAATC AACTGTTACA TGGAATGAAA CTTCCAGACA CCGTACTTTA	180
GTTCATATT TAAAACATGT TGAGCTACAG CACCAATTAG AGCAATTAAAG CTCTAAGCCA	240
TCCGCAAAAAA TGACCTCTTA TCAAAAGGAG CAATTAAAGG TACTCTCTAA TCCTGACCTG	300
TTGGAGTTTG CTTCCGGTCT GGTTCGCTTT GAAGCTCGAA TTAAAACCGC ATATTTGAAG	360
TCTTCGGGG TTCCTCTTAA TCTTTTGAT GCAATCCGCT TTGCTTCTGA CTATAATAGT	420
CAGGGTAAAG ACCTGATTT TGATTATGG TCATTCTCGT TTTCTGAACG GTTAAAGCA	480
TTTGAGGGGG ATTCAATGAA TATTTATGAC GATTCCGAG TATTGGACGC TATCCAGTCT	540
AAACATTTTA CTATTACCCC CTCTGGCAAA ACTTCTTTG CAAAAGCCTC TCGCTATTAA	600
GGTTTTATC GTCGCTGGT AAACGGAGGGT TATGATAGTG TTGCTCTTAC TATGCCCTGGT	660
AATTCCCTTT GGCCTTATGT ATCTGCATTA GTTGAATGTG GTATTCCTAA ATCTCAACTG	720
ATGAATCTTT CTACCTGAA TAATGTTGTT CCGTTAGTCG GTTTTATTAA CGTAGATTTT	780
TCTTCCCAAC GTCCTGACTG GTATAATGAG CCAGTTCTTA AAATCGCATA AGGTAATTCA	840
CAATGATTAA AGTTGAAATT AAACCATCTC AAGCCCAATT TACTACTCGT TCTGGTGT	900
CTCGTCAGGG CAAGCCTTAT TCACTGAATG AGCAGCTTIG TTACGTTGAT TTGGGTAATG	960
AATATCCGGT TCTTGTCAAG ATTACTCTTG ATGAAGGTCA GCCAGCCTAT GCGCCTGGTC	1020
TGTACACCGT TCATCTGTCC TCTTCAAAG TTGGTCAGTT CGGTTCCCTT ATGATTGACC	1080
GTCTGCGCCT CGTTCCGGCT AAGTAACATG GAGCAGGTGCG CGGATTCGA CACAATTAT	1140
CAGGCGATGA TACAAATCTC CGTTGTACTT TGTTTGCCTC TTGGTATAAT CGCTGGGGT	1200
CAAAGATGAG TGTGTTAGTG TATTCTTCG CCTCTTCGT TTTAGGTTGG TGCCCTGGTA	1260

GTGGCATTAC GTATTTAACC CGTTTAATGG AAACTTCCTC ATGAAAAAGT CTTTAGTCCT	1320
CAAAGCCTCT GTAGCCGTTG CTACCCCTCGT TCCGATGCTG TCTTCGCTG CTGAGGGTGA	1380
CGATCCCGCA AAAGCGGCCT TAAACTCCCT GCAAGCCTCA GCGACCGAAT ATATCGGTTA	1440
TGCGTGGGCG ATGGTGTGTC TGATTGTCGG CGCAACTATC GGTATCAAGC TGTTTAAGAA	1500
ATTCACCTCG AAAGCAAGCT GATAAACCGA TACAATTAAA GGCTCCTTT GGAGCCTTT	1560
TTTTGGAGA TTTCAACGT GAAAAAATTA TTATTCGAA TTCTTTAGT TGTTCCCTTC	1620
TATTCTCACT CCGCTGAAAC TGTGAAAGT TGTTAGCAA AACCCCATAC AGAAAATTCA	1680
TTTACTAACG TCTGGAAAGA CCACAAAAGT TTAGATCGTT ACGCTAACTA TGAGGGTTGT	1740
CTGTGGAATG CTACAGGCGT TGTAGTTGT ACTGGTGACG AAACTCAGTG TTACGGTACA	1800
TGGGTTCTA TTGGGCTTGC TATCCCTGAA AATGAGGGTG GTGGCTCTGA GGGTGGCGGT	1860
TCTGAGGGTG GCGGTTCTGA GGGTGGCGGT ACTAACCTC CTGAGTACGG TGATACACCT	1920
ATTCCGGGCT ATACTTATAT CAACCCCTCTC GACGGCACTT ATCCGCCCTGG TACTGAGCAA	1980
AACCCCGCTA ATCCTAATCC TTCTCTTGAG GAGTCTCAGC CTCTTAATAC TTTCATGTTT	2040
CAGAATAATA GGITCCGAAA TAGGCAGGGG GCATTAACIG TTTATACGGG CACTGTTACT	2100
CAAGGCACIG ACCCCCCTAA AACTTATTAAC CAGTACACTC CTGTATCATC AAAAGCCATG	2160
TATGACGCTT ACTGGAACGG TAAATTCAAGA GACTGCCCTT TCCATTCTGG CTITAATGAA	2220
GATCCATTG TTTGTGAATA TCAAGGCCAA TCGCTCTGACCC TGCCCTCAACC TCCTGTCAAT	2280
GCTGGCGGGG GCTCTGGTGG TGTTCTGGT GGCGGCTCTG AGGGTGGTGG CTCTGAGGGT	2340
GGCGGTTCTG AGGGTGGCGG CTCTGAGGGG GGCGGTTCCG GTGGTGGCTC TGTTCCGGT	2400
GATTTTGATT ATGAAAAGAT GGCAAAACGCT AATAAGGGGG CTATGACCGA AAATGCCGAT	2460
AAAAACGCCG TACAGTCTGA CGCTAAAGGC AAACCTGATT CTGTGCTAC TGATTACGGT	2520
GCTGCTATCG ATGGTTTCAT TGTTGACGTT TCCGGCTTG CTAATGGTAA TGTTGCTACT	2580
GGTGATTTG CTGGCTCTAA TTCCCAAATG GCTCAAGTGG GTGACGGTGA TAATTACACCT	2640
TTAATGAATA ATTTCCGTCA ATATTTACCT TCCCTCCCTC AATCGGTTGA ATGTCGCCCT	2700
TTTGTCTTTA GCGCTGGTAA ACCATATGAA TTTTCTATTG ATTGTGACAA AATAAAACTTA	2760
TTCCGTGGTG TCTTTGCTGG TCTTTTATAT GTTGGCACCT TTATGTATGT ATTTTCTACG	2820
TTTGCTAACAA TACTGCGTAA TAAGGAGTCT TAATCATGCC AGTCTTTTG GGTATTCCGT	2880
TATTATTGGG TTCCCTCGGT TTCCCTCTGG TAACTTGTGTT CGGCTATCTG CTTACTTTTC	2940
TTAAAAAGGG CTTCGGTAAG ATAGCTATTG CCTGTTCTT GCTCTTATTA TTGGGCTTAA	3000
CTCAATTCTT GTGGGTTATC TCTCTGATAT TAGCGCTCAA TTACCCCTCG ACTTTGTTCA	3060
GGGTGTTCAAG TTAATTCTCC CGTCTAACATC GCTTCCCTGT TTTTATGTTA TTCTCTCTGT	3120
AAAGGCTGCT ATTTTCAATT TTGACGTTAA ACAAAAATC GTTCTTATT TGGATTGGGA	3180
TAAATAATAT GGCTGTTAT TTTGTAACIG GCAAATTAGG CTCTGGAAAG ACGCTCGTTA	3240
GGCGTTGGTAA GATTCAAGGAT AAAATTGTAG CTGGGTGCAA AATAGCAACT AATCTGATT	3300

TAAGGCTTCA	AAACCTCCCC	CAAGTCGGGA	GGTTCGCTAA	AACGCCCTGGC	GTTCTTAGAA	3360
TACCGGATAA	GCCTTCTATA	TCTGATTGCG	TTGCTATTGG	GCGCGGTAAT	GATTCCCTACG	3420
ATGAAAATAA	AAACGGCTTG	CTTGTCTCG	ATGAGTGCAG	TACTTGGTTT	AATAACCGTT	3480
CTTGGAAATGA	TAAGGAAAGA	CAGCCGATTA	TTGATTGGTT	TCTACATGCT	CGTAAATTAG	3540
GATGGGATAT	TATTTTCTT	GTTCAGGACT	TATCTATTGT	TGATAAACAG	GCGCGTTCTG	3600
CATTAGCTGA	ACATGTTGTT	TATTGTCGTC	GTCTGGACAG	AATTACTTTA	CCTTTTGTGG	3660
GTACTTTATA	TTCTCTTATT	ACTGGCTCGA	AAATGCCCTCT	GCCTAAATTAA	CATGTTGGCG	3720
TTGTTAAATA	TGGCGATTCT	CAATTAAGCC	CTACTGTTGA	GCGTTGGCTT	TATACTGGTA	3780
AGAATTGTA	TAACCGATAT	GATACTAAAC	AGGCTTTTTC	TAGTAATTAT	GATTCCGGTG	3840
TTTATTCTTA	TTAACGGCCT	TATTTATCAC	ACGGTCGGTA	TTTCAAACCA	TTAAATTAG	3900
GTCAGAAGAT	GAAGCTTACT	AAAATATATT	TGAAAAAGTT	TTCACGCGTT	CTTGTCTTG	3960
CGATTGGATT	TGCATCAGCA	TTTACATATA	GTTATATAAC	CCAACCTAAAG	CCGGAGGTTA	4020
AAAAGGTAGT	CTCTCAGACC	TATGATTITG	ATAAATTAC	TATTGACTCT	TCTCAGCGTC	4080
TTAATCTAAG	CTATCGCTAT	GTTTTCAAGG	ATTCTAAGGG	AAAATTAAATT	AA TAGCCAGG	4140
ATTTACAGAA	GCAAGGTTAT	TCACTCACAT	ATATTGATT	ATGACTGTT	TCCATTAAAA	4200
AAGGTAATT	AAATGAAATT	GTAAATGTA	ATTAATT	TTTCTTGAT	GTITGTTCA	4260
TCATCTTCTT	TTGCTCAGGT	AATTGAAATG	AATAATTCGC	CTCTGCGCGA	TTTTGTAACT	4320
TGGTATTCAA	AGCAATCAGG	CGAATCCGTT	ATTGTTCTC	CCGATGTAAA	AGGTACTGTT	4380
ACTGTATATT	CATCTGACGT	AAAACCTGAA	AATCTACGCA	ATTCTTTAT	TTCTGTTTA	4440
CGTGCTAATA	ATTTGATAT	GGTTGGTTCA	ATTCCCTCCA	TAATTCAAGAA	GTATAATCCA	4500
AACAATCAGG	ATTATATTGA	TGAATTGCCA	TCATCTGATA	ATCAGGAATA	TGATGATAAT	4560
TCCGCTCCCT	CTGGTGGTTT	CTTGTGTCGG	CAAAATGATA	ATGTTACTCA	AACTTTAAA	4620
ATTAATAACG	TTCGGGCAAA	GGATTTAATA	CGAGTTGTCG	AATTGTTGT	AAAGTCTAAT	4680
ACTTCTAAAT	CCTCAAATGT	ATTATCTATT	GACGGCTCTA	ATCTATTAGT	TGTTAGTGCA	4740
CCTAAAGATA	TTTAGATAA	CCTTCCTCAA	TTCCCTTCTA	CTGTTGATTT	GCCAACTGAC	4800
CAGATATTGA	TTGAGGGTTT	GATATTGAG	GTTCAGCAAG	GTGATGCTTT	AGATTTTCA	4860
TTTGCTGCTG	GCTCTCAGGG	TGGCACTGTT	GCAGGCCGTG	TTAATACTGA	CCGGCTCACC	4920
TCTGTTTAT	CTTCTGCTGG	TGGTTCGTT	GGTATTTTTA	ATGGCGATGT	TTTAGGGCTA	4980
TCAGTTCGGG	CATTAAGAC	TAATAGCCAT	TCAAAATAT	TGTCTGTGCC	ACGTATTCTT	5040
ACGCTTTCAAG	GTCAGAAGGG	TTCTATCTCT	GTGCGCCAGA	ATGTCCTTT	TATTACTGGT	5100
CGTGTGACTG	GTGAATCTGC	CAATGTAAAT	AATCCATTTC	AGACGATTGA	GCGTAAAT	5160
GTAGGTATT	CCATGAGCGT	TTTCTCTGTT	GCAATGGCTG	GCGGTAATAT	TGTTCTGGAT	5220
ATTACCAGCA	AGGCCGATAG	TTGAGTTCT	TCTACTCAGG	CAAGTGTATGT	TATTACTAAT	5280
CAAAGAAGTA	TTGCTACAAAC	GGTTAATTG	CGTGATGGAC	AGACTCTTT	ACTCGGTGGC	5340

CTCACTGATT ATAAAAACAC TTCTCAAGA-	TCTGGCGTAC CGTTCCGTGTC TAAAATCCCT	5400
TTAACCGGGC TCCTGTTAG CTCCCGCTCT GATTCCAACG AGGAAAGCAC GTTATACTGTG	5460	
CTCGTCAAAG CAACCATAGT ACGCGCCCTG TAGCGGGCGA TTAAGCGCGG CGGGTGTGGT	5520	
GGTGTACCGGC AGCGTGACCG CTACACTTGC CAGGGCCCTA CGCCCGCTC CTTCGCTTT	5580	
CTTCCCTTCC TTTCTCGCCA CGTTCGCCGG CTTTCCCCGT CAAGCTCTAA ATCGGGGGCT	5640	
CCCTTAGGG TTCCGATTAA GTGCTTACG GCACCTCCGAC CCCAAAAAAC TTGATTTGGG	5700	
TGATGGTTCA CGTAGTGGGC CATCGCCCTG ATAGACGGTT TTTCGCCCTT TGACGTTGGA	5760	
GTCCACGTTTCTTAAATAGTG GACTCTTGTG CCAAACGTGAA ACAACACTCA ACCCTATCTC	5820	
GGGCTATTCT TTTGATTTAT AAGGGATTTT GCGGATTTCG GAACCACCAT CAAACAGGAT	5880	
TTTCCGCTGC TGGGGCAAC CAGCGTGGAC CGCTTGTGTC AACTCTCTCA GGGCCAGGCG	5940	
GTGAAGGGCA ATCAGCTGTT GCCCGTCTCG CTGGTAAAAA GAAAAACAC CCTGGGCC	6000	
AATAACGCAAA CGGCCTCTCC CGCGCGCTTG GCCGATTCA TAATGCAGCT GGCACGACAG	6060	
GTTTCCCGAC TGGAAAGCGG GCAGTGAGCG CAACGCAATT AATGTGAGTT AGCTCA	6120	
TTAGGCACCC CAGGCTTAC ACTTTATGCT TCCGGCTCGT ATGTTGTGTC GAATTGAG	6180	
CGGATAACAA TTTCACACGC CAAGGAGACA GTCATAATGA AATACCTATT GCCTACGGCA	6240	
GCCGCTGGAT TGTTATTACT CGCTGCCAA CCAGCCATGG CCGAGCTCTT CCCGCCATCT	6300	
GATGAGCAGT TGAAATCTGG AACTGCCCTCT GTTGTGTGCC TGCTGAATAA CTTCTATCCC	6360	
AGAGAGGCCA AACTACAGTG GAAGGTGGAT AACGCCCTCC AATCGGGTAA CTCCGAGGAG	6420	
ACTGTCACAG ACCAGGACAG CAAGGACAGC ACCTACAGCC TCAGCGACAC CCTGACGCTG	6480	
AGCAAAGCAG ACTACGAGAA ACACAAAGTC TACGCCCTGC AAGTCACCCA TCAGGGCCTG	6540	
AGCTCGCCCC TCACAAAGAG CTTCAACAGG GGAGAGTGTGTT CTAGAACGGC TCACCTGGCA	6600	
CTGGCCGTGCG TTTTACAACG TCGTGAATGGGAA GAAAACCTTG CGCTTACCCA AGCTTAATCG	6660	
CCTTGCAGAA TTCCCTTCG CCAGCTGGCG TAATAGCGAA GAGGCCCGCA CCGATGCC	6720	
TTCCCAACAG TTGGCAGCC TGAATGGCGA ATGGCCCTTT GCCTGGTTTC CGGCACCAAGA	6780	
AGCGGTGCCG GAAACCTGGC TGGAGTGCAGA TCTTCCTGAG GCCGATACGG TCGTCGTCCC	6840	
CTCAAACCTGG CAGATGCACG GTTACGATGC GCCCATCTAC ACCAACGTAA CCTATCCCAT	6900	
TACGGTCAAT CCGCCGTTTG TTCCACCGA GAATCCGACG GGTGTTACT CGCTCACATT	6960	
TAATGTTGAT GAAAGCTGGC TACAGGAAGG CCAGACGGCA ATTATTTTG ATGGCGTTCC	7020	
TATTGGTTAA AAAATGAGCT GATTTAACAA AAATTTAACG CGAATTTAA CAAAATATTA	7080	
ACGTTTACAA TTAAATATT TGCTTATACA ATCTTCCTGT TTTGGGGCT TTTCTGATTA	7140	
TCAACCGGGG TACATATGAT TGACATGCTA GTTTTACGAT TACCGTTCAT CGATTCTCTT	7200	
GTTTGCTCCA GACTCTCAGG CAATGACCTG ATAGCCTTTC TAGATCTCTC AAAAATAGCT	7260	
ACCCCTCTCGG GCATTAATTT ATCAGCTAGA ACGGTTGAAT ATCATATTGA TGGTGATTTG	7320	
ACTGTCTCCG GCCTTTCTCA CCCTTTGAA TCTTACCTA CACATTACTC AGGCATTGCA	7380	

TTTAAAATAT ATGAGGGTTC TAAAAATTTC TATCCTTGCG TTGAAATAAA GGCTTCTCCC	7440
GCAAAAGTAT TACAGGGTCA TAATGTTTT GGTACAACCG ATTTAGCTTT ATGCTCTGAG	7500
GCTTATTGCC TTAATTTCGC TAATTCTTTG CTTGCCTGT ATGATTATT GGATGTT	7557

(2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 8118 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: both
 - (D) TOPOLOGY: circular

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

AATGCTACTA CTATTAGTAG AATTGATGCC ACCTTTTCAG CTCGGGGCCC AAATGAAAAT	60
ATAGCTAACAC AGGTTATTGA CCATTGCGA AATGTATCTA ATGGTCAAAC TAAATCTACT	120
CGTTCCGAGA ATTGGGAATC AACTGTTACA TGGAATGAAA CTTCCAGACA CCGTACTTTA	180
TTTGCATATT TAAAACATGT TGAGCTACAG CACCAGATTC AGCAATTAAG CTCTAAGCCA	240
TCTGAAAAAA TGACCTCTTA TCAAAAGGAG CAATTAAAGG TACTCTCTAA TCCTGACCTG	300
TTGGAGTTG CTTCCGGTCT GGTTCGCTTT GAAGCTCGAA TTAAAACCGG ATATTTGAAG	360
TCTTCGGGC TTCCTCTAA TCTTTTGAT GCAATCCGCT TTGCTCTGA CTATAATAGT	420
CAGGGTAAAG ACCTGATTT TGATTTATGG TCATTCTCGT TTTCTGAAC GTTTAAAGCA	480
TTTGAGGGGG ATTCAATGAA TATTTATGAC GATTCCGCAG TATTGGACGC TATCCAGTCT	540
AAACATTTTA CTATTACCCC CTCTGGCAAA ACTTCTTTG CAAAAGCCTC TCGCTATTIT	600
GGTTTTTATC GTCGTCTGGT AAACGAGGGT TATGATAGTG TTGCTCTTAC TATGCCTCGT	660
AATTCCCTTT GGCGTTATGT ATCTGCATTA GTTGAATGTG GTATTCTAA ATCTCAACTG	720
ATGAATCTTT CTACCTGTAA TAATGTTGTT CCGTTAGTTC GTTTTATTAA CGTAGATTIT	780
TCTTCCCAAC GTCCCTACTG GTATAATGAG CCAGTTCTTA AAATGGATA AGGTAATTCA	840
CAATGATTAA AGTTGAAATT AAACCATCTC AAGCCCAATT TACTACTCGT TCTGGTGTIT	900
CTCGTCAGGG CAAGCCITAT TCACTGAATG AGCAGCTTTG TTACGTTGAT TTGGGTAATG	960
AATATCCGGT TCTTGTCAAG ATTACTCTTG ATGAAGGTCA GCCAGCCTAT GGGCCTGGTC	1020
TGTACACCGT TCATCTGTCC TCTTCAAAG TTGGTCAGTT CGGTTCCCTT ATGATTGACC	1080
GTCTGGCCCT CGTTCCGGCT AAGTAACATG GAGCAGGTGCG CGGATTTGGA CACAATTAT	1140
CAGGCATGA TACAAATCTC CGTTGTACTT TGTTTCGCGC TTGGTATAAT CGCTGGGGGT	1200
CAAAGATGAG TGTTTGTG TATTCTTCG CCTCTTCTCGT TTTAGGTTGG TGCCTTCGTA	1260
GTGGCATTAC GTATTTTACCG CGTTTAATGG AAACCTCCCTC ATGAAAAAGT CTTTAGTCCT	1320
CAAAGCCTCT GTAGCCGTG CTACCCCTCGT TCCGATGCTG TCTTTCGCTG CTGAGGGTGA	1380
CGATCCCCGA AAAGCCGCCT TTAACTCCCT GCAAGCCTCA GCGACCGAAT ATATCGGTTA	1440
TGCGTGGGGCG ATGGTTGTTG TCATTGTCCG CGCAACTATC GGTATCAAGC TGTTTAAGAA	1500

ATTCACCTGG AAAGCAAGCT GATAAACCGA TACAATTAAA GGTCCTTTT GGAGCCTTT	1560
TTTTGGAGA TTTTCAACGT GAAAAAAATTA TTATTCCCAA TTCCCTTAGT TGTCCTTTTC	1620
TATTCTCACT CCGCTGAAAC TGTTGAAAGT TGTTIAGCAA AACCCCATAAC AGAAAATTCA	1680
TTTACTAACG TCTGGAAAGA CGACAAAAGT TTAGATCGTT ACGCTAACTA TGAGGGTTGT	1740
CTGTGGAATG CTACAGGCCGT TGTAGTTGT ACTGGTGACG AAACTCAGTG TTACGGTACA	1800
TGGGTTCCCTA TTGGGCTTGC TATCCCTGAA AATGAGGGTG GTGGCTCTGA GGGTGGCGGT	1860
TCTGAGGGTG GCGGTTCTGA GGGTGGCGGT ACTAAACCTC CTGACTACGG TGATACACCT	1920
ATTCCGGGCT ATACTTATAT CAACCCCTCTC GACGGCACTT ATCCGCTTGG TACTGAGCAA	1980
AACCCCGCTA ATCCCTAATCC TTCTCTTGAG GAGTCTCAGC CTCTTAATAC TTTCATGTT	2040
CAGAATAATA GGTTCCGAAA TAGGCAGGGG GCATTAACTG TTTATACGGG CACTGTTACT	2100
CAAGGCACTG ACCCCGTTAA AACTTATTAC CAGTACACTC CTGTATCATC AAAAGCCATG	2160
TATGACGCTT ACTGGAACGG TAAATTAGA GACTGCGCTT TCCATTCTGG CTTTAATGAA	2220
GATCCATTG TTTGTGAATA TCAAGGCCAA TCGTCTGACC TGCCCTCAACC TCCTGTCAAT	2280
GCTGGCGGGG GCTCTGGTGG TGTTCTGGT GGCGGCTCTG AGGGTGGTGG CTCTGAGGGT	2340
GGCGGTTCTG ACGGTGGCGG CTCTGAGGGGA GGCGGTTCCG GTGGTGGCTC TGTTCCGGT	2400
GATTTTGATT ATGAAAAGAT GGCAAACGCT AATAAGGGGG CTATGACCGA AAATGCCGAT	2460
GAAAACGCCG TACAGTCTGA CGCTAAAGGC AAACTTGATT CTGTCGCTAC TGATTACGGT	2520
GCTGCTATCG ATGGTTTCA TGGTGACGTT TCCGGCTTG CTAATGGTAA TGGTGCTACT	2580
GGTGATTTG CTGGCTCTAA TTCCCAAATG GCTCAAGTCG GTGACGGTGA TAATTCACT	2640
TTAATGAATA ATTTCCGTCA ATATTTACCT TCCCTCCCTC AATCGGTTGA ATGTCGCCCT	2700
TTTGTCTTAA GCGCTGGTAA ACCATATGAA TTTTCTATTG ATTGTGACAA AATAAAACTTA	2760
TTCCGTGGTG TCTTTGGTGT TCTTTATAT GTGCCACCT TTATGTATGT ATTTTCTACG	2820
TTTGCTAACAA TACTGCGTAA TAAGGAGTCT TAATCATGCC AGTTCTTTG GGTATTCCGT	2880
TATTATTGCG TTCCCTCGGT TTCCCTCTGG TAACTTTGTT CGGCTATCTG CTTACTTTTC	2940
TTAAAAAGGG CTTCGGTAAG ATAGCTATTG CTATTCATT GTTTCTTGCT CTTATTATTG	3000
GGCTTAACTC AATTCTGTG GGTTATCTCT CTGATATTAG CGCTCAATTAA CCCTCTGACT	3060
TTGTTCAAGG TGTTCAAGTAA ATTCTCCGT CTAATCGCT TCCCTGTTT TATGTTATTC	3120
TCTCTGTAAA GGCTGCTATT TTGATTTTG ACGTTAAACA AAAAATCGTT TCTTATTTGG	3180
ATTGGGATAAA ATAATATGGC TGTTTATTTT GTAATGGCA AATTAGGCTC TGGAAAGACG	3240
CTCGTTAGCG TTGGTAAGAT TCAGGATAAA ATTGTAGCTG GGTGCAAAAT AGCAACTAAT	3300
CTTGATTTAA GGCTCAAAA CCTCCCGCAA GTCGGGAGGT TCGCTAAAC GCCTCGCGTT	3360
CTTAGAACAC CGGATAAGCC TTCTATATCT GATTTGCTTG CTATTGGGG CGGTAATGAT	3420
TCCTACGATG AAAATAAAA CGGCTTGCTT GTTCTCGATG AGTGGGGTAC TTGGTTTAAT	3480
ACCCGTTCTT GGAATGATAA CGAAAGACAG CCGATTATTG ATTGGTTCT ACATGCTCGT	3540

AAATTAGGAT GGGATATTAT TTTCTTGTG CAGGACTTAT CTATTGTTGA TAAACAGGCG	3600
CGTTCTGCAT TAGCTGAACA TGTTGTTAT TGTCGTCGTC TGGACAGAAT TACTTACCT	3660
TTTGTGGTA CTTTATATTG TCTTATTACT GGCTCGAAAA TGCCCTGCC TAAATTACAT	3720
GTTGGCGTTG TTAAATATGG CGATTCTGAA TTAAGCCTA CTGTTGAGCG TTGGCTTAT	3780
ACTGGTAAGA ATTTGTATAA CGCATATGAT ACTAACAGG CTTTTCTAG TAATTATGAT	3840
TCCGGTGTGTT ATTCTTATTT AACGCCTTAT TTATCACACG GTCGGTATTG CAAACCATT	3900
AATTTAGGTC AGAAGATGAA CCTTACTAAA ATATATTGAA AAAAGTTTC ACGCGTTCTT	3960
TGTCTTGCAG TTGGATTGTC ATCAGCATT ACATATAGTT ATATAACCCA ACCTAACCG	4020
GAGGTTAAAA AGGTAGCTC TCAGACCTAT GATTTGATA AATTGACTAT TGACTCTCT	4080
CAGCGTCTTA ATCTAAGCTA TCGCTATGTT TTCAAGGATT CTAAGGGAAA ATTAATTAAT	4140
AGCGACGATT TACAGAACCA AGGTTATTCA CTCACATATA TTGATTTATG TACTGTTCC	4200
ATTAAAAAAG GTAATTCAAA TGAAATTGTT AAATGTAATT AATTTGTTT TCTTGATGTT	4260
TGTTTCATCA TCTTCTTTG CTCAGGTAAT TGAAATGAAT AATTGCCTC TGCGGATT	4320
TGTAACTTGG TATTCAAAGC AATCAGGCGA ATCCGTTATT GTTCTCCCCG ATGTAAGG	4380
TACTGTTACT GTATATTCACT CTGACGTTAA ACCTGAAAAT CTACGGAATT TCTTTATTC	4440
TGTTTTACGT GCTAATAATT TTGATATGGT TGTTCAATT CCTTCCATAA TTCAGAAGTA	4500
TAATCCAAAC AATCAGGATT ATATTGATGA ATTGCCATCA TCTGATAATC AGGAATATGA	4560
TGATAATTCC GCTCCCTCTG GTGGTTCTT TGTTCCGAA AATGATAATG TTACTCAAAC	4620
TTTTAAATT AATAACGTTG GGGCAAAGGA TTAAATACGA GTTGTGAAT TGTTGTAAA	4680
GTCTAATACT TCTAAATCCT CAAATGTATT ATCTATTGAC GGCTCTAATC TATTAGTTG	4740
TAGTGCACCT AAAGATATT TAGATAACCT TCCTCAATC CTTCTACTG TTGATTTGCC	4800
AACTGACCAAG ATATTGATTG AGGTTTGAT ATTTGAGGTT CAGCAAGGTG ATGCTTAA	4860
TTTTCAATT GCTGCTGGCT CTCAGCGTGG CACTGTTGCA GGCGGTGTTA ATACTGACCG	4920
CCTCACCTCT GTTTATCTT CTGCTGGTGG TTGCTCGGT ATTTTAATG GCGATGTTT	4980
AGGGCTATCA GTTCGCGCAT TAAAGACTAA TAGCCATTCA AAAATATTGT CTGTGCCACG	5040
TATTCTTACG CTTTCAGGTC AGAAGGGTC TATCTCTGTT GGCCAGAATG TCCCTTTAT	5100
TACTGGTCGT GTGACTGGTG AATCTGCCAA TGAAATAAT CCATTTAGA CGATTGAGCG	5160
TCAAAATGTA GGTATTCCG TGAGGTTTT TCCTGTTGCA ATGGCTGGCG GTAATATTGT	5220
TCTGGATATT ACCAGCAAGG CCGATAGTTT GAGTTCTCT ACTCAGGCAA GTGATGTTAT	5280
TACTAATCAA AGAAGTATTG CTACAACGGT TAAATTGCGT GATGGACAGA CTCTTTACT	5340
CGGTGGCCTC ACTGATTATA AAAACACTTC TCAAGATTCT GGCGTACCGT TCCCTGTCAA	5400
AATCCCTTA ATCGGCCTCC TGTTAGCTC CCGCTCTGAT TCCAACGAGG AAAGCACGTT	5460
ATACGTGCTC GTCAAAGCAA CCATAGTACG CGCCCTGTAG CGGGCGATTA AGCGCGGGCG	5520
GTGTGGTGGT TACGCGCAGC GTGACCGCTA CACTTGGCAG CGCCCTAGCG CCCGCTCCTT	5580

TCGGTTTCTT	CCCTTCCTT	CTCGCCACGT	TCGGGGCTT	TCCCCGTCAA	GCTCTAAATC	5640
GGGGGCTCCC	TTTAGGGTTC	CGATTTAGTG	CTTTAUGGCA	CCTCGACCCC	AAAAAACTTG	5700
ATTTGGGTGA	TGGTTCACGT	AGTGGGCCAT	CGCCCTGATA	GACGGTTTTT	CGCCCTTGA	5760
CGTTGGAGTC	CACGTTCTT	AATAGTGGAC	TCTTGTCCA	AACTGGAACA	ACACTCAACC	5820
CTATCTCGGG	CTATTCTTT	GATTTATAAG	GGATTTGCC	GATTTCGGAA	CCACCATCAA	5880
ACAGGATTTT	CGCCTGCTGG	GGCAAACCCAG	CGTGGACCGC	TTGCTGCAAC	TCTCTCAGGG	5940
CCAGGCGGTG	AAAGGCAATC	AGCTGTTGCC	CGTCTCGCTG	GTGAAAAGAA	AAACCACCCCT	6000
GGCGCCCAAT	ACGCAAACCG	CCTCTCCCCG	CGCGTTGCC	GATTCACTAA	TGCAGCTGGC	6060
ACGACAGGTT	TCCCAGTGG	AAAGGGGCA	GTGAGCGAA	CGCAATTAAT	GTGAGTTAGC	6120
TCACTCATTA	GGCACCCAG	GCTTACACT	TTATGCTTCC	GGCTCGTATG	TTGTGTGGAA	6180
TTGTGAGCGG	ATAACAATTT	CACACGCCAA	GGAGACAGTC	ATAATGAAAT	ACCTATTGCC	6240
TACGGCAGCC	GCTGGATTGT	TATTACTCGC	TGCCCAACCA	GCCATGGCCG	AGCTCTTCCC	6300
GCCATCTGAT	GAGCAGTTGA	AATCTGGAAC	TGCCCTGTT	GTGTGCTGCC	TGAATAACTT	6360
CTATCCCAGA	GAGGCCAAAG	TACAGTGGAA	GGTGGATAAC	GCCCTCCAAT	CGGGTAACTC	6420
CCAGGAGAGT	GTCACAGAGC	AGGACAGCAA	GGACAGCACC	TACAGCCTGA	GCAGCACCCCT	6480
GACGCTGAGC	AAACCAAGACT	ACGAGAAACA	CAAAGTCTAC	GCCTGCGAAG	TCACCCATCA	6540
GGGCCTGAGC	TGGCCCGTCA	CAAAGAGCTT	CAACAGGGGA	GAGTGTCTA	GAACGCGTCA	6600
CTTGGCACTG	GGCGTCGTTT	TACAACGTG	TGACTGGAA	AAACCTGGCG	TTACCCAAGC	6660
TTTGATCATG	GAGAAAATAA	AGTGAACAA	AGCACTATTG	CACTGGCACT	CTTACCGTTA	6720
CTGTTTACCC	CTGTGGAAA	AGCCGCCTCC	ACCAAGGGCC	CATOGGTCTT	CCCCCTGGCA	6780
CCCTCCTCCA	AGACCACTC	TGGGGCACA	GGGGCCCTGG	GCTGCCCTGGT	CAAGACTAAT	6840
TCCCCGAACC	GGTGACGGTG	TGGTGGAACT	CAGGCGCCCT	GACCAGCGGC	GTGCACACCT	6900
TCCCCGCTGT	CCTACAGTCC	TCAGGACTCT	ACTCCCTCAG	CAGCGTGGTG	ACCGTGCCCT	6960
CCAGCAGCTT	GGGCACCCAG	ACCTACATCT	GCAACGTGAA	TCACAAGCCC	AGGAACACCAA	7020
AGGTGGACAA	GAAAGCAGAG	CCCAAATCTT	GTACTAGTGG	ATCCTACCCG	TACGACGTT	7080
CGGACTACGC	TTCTTACGGCT	GAAGGCCATG	ACCCCTGCTAA	GGCTGCATTG	AATAGTTAC	7140
AGGCAAGTGC	TACTGAGTAC	ATTGGCTACG	CTTGGCTAT	GGTAGTAGTT	ATAGTTGGTG	7200
CTACCATAGG	GATTAATTAA	TTCAAAAAGT	TTACGAGCAA	GGCTCTTAA	CCAATAGCGA	7260
AGAGGCCCGC	ACCGATCGCC	CTTCCAACA	GTGCGCAGC	CTGAATGGCG	AATGGCGCTT	7320
TGCCCTGGTTT	CCGGCACCAAG	AAGCGGTGCC	GGAAAGCTGG	CTGGAGTGCG	ATCTTCCCTGA	7380
GGCCGATACCG	GTCGTGCTCC	CCTCAAACGT	GCAGATGCAC	GGTTACGATG	CGCCCATCTA	7440
CACCAACGTA	ACCTATCCC	TTACGGTCAA	TCCGGCTTT	GTTCACCGG	AGAATCCGAC	7500
GGGTGTTAC	TCGCTGACAT	TTAATGTTGA	TGAAAGCTGG	CTACAGGAAG	GCCAGACGCC	7560
AATTATTTTT	GATGGCGTTC	CTATTGGTTA	AAAAATGAGC	TGATTTAAC	AAAATTTAAC	7620

62

GGGAATTTA ACAAAATATT AACGTTTACA ATTAAATAT TTGCTTATAC AATCTTCCTG	7680
TTTTGGGGC TTTTCTGATT ATCAACCGGG GTACATATGA TTGACATGCT AGTTTTACGA	7740
TTACCGTTCA TCGATTCTCT TGTTTGCTCC AGACTCTCAG GCAATGACCT GATAGCCTTT	7800
GTAGATCTCT CAAAAATAGC TACCCCTCTCC GGCATTAATT TATCAGCTAG AACGGTTGAA	7860
TATCATATTG ATGGTGTATT GACTGTCTCC GGCCTTTCTC ACCCTTTGA ATCTTTACCT	7920
ACACATTACT CAGGCATTGC ATTAAAATA TATGAGGGTT CTAAAAATT TTATCCTTGC	7980
GTTGAAATAA AGGCTTCTCC CGCAAAAGTA TTACAGGGTC ATAATGTTT TGGTACAACC	8040
GATTAGCTT TATGCTCTGA GGCTTTATTG CTTAATTG CTAATTCTTT GCCTTGCCTG	8100
TATGATTTAT TGGACGTT	8118

(2) INFORMATION FOR SEQ ID NO:6:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: misc_difference
- (B) LOCATION: replace(5, "")
- (D) OTHER INFORMATION: /note= "S REPRESENTS EQUAL MIXTURE OF G AND C"

(ix) FEATURE:

- (A) NAME/KEY: misc_difference
- (B) LOCATION: replace(6, "")
- (D) OTHER INFORMATION: /note= "M REPRESENTS EQUAL MIXTURE OF A AND C"

(ix) FEATURE:

- (A) NAME/KEY: misc_difference
- (B) LOCATION: replace(8, "")
- (D) OTHER INFORMATION: /note= "R REPRESENTS EQUAL MIXTURE OF A AND G"

(ix) FEATURE:

- (A) NAME/KEY: misc_difference
- (B) LOCATION: replace(11, "")
- (D) OTHER INFORMATION: /note= "K REPRESENTS EQUAL MIXTURE OF G AND T"

(ix) FEATURE:

- (A) NAME/KEY: misc_difference
- (B) LOCATION: replace(20, "")
- (D) OTHER INFORMATION: /note= "W REPRESENTS EQUAL MIXTURE OF A AND T"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

AGGTSMARCT KCTCGAGTCW GG

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

AGGTCCAGCT GCTCGAGTCT GG

22

(2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

AGGTCCAGCT GCTCGAGTCA GG

22

(2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

AGGTCCAGCT TCTCGAGTCT GG

22

(2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

AGGTCCAGCT TCTCGAGTCA GG

22

(2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

64

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

AGGTCCAACT GCTCGAGTCT GG

22

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

AGGTCCAACT GCTCGAGTCA GG

22

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

AGGTCCAACT TCTCGAGTCT GG

22

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

AGGTCCAACT TCTCGAGTCA GG

22

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: misc_difference
- (B) LOCATION: replace(5..6, "")
- (D) OTHER INFORMATION: /note= "N-INOSINE"

(ix) FEATURE:

- (A) NAME/KEY: misc_difference
- (B) LOCATION: replace(8, "")
- (D) OTHER INFORMATION: /note= "N-INOSINE"

(ix) FEATURE:

- (A) NAME/KEY: misc_difference
- (B) LOCATION: replace(11, "")
- (D) OTHER INFORMATION: /note= "N=INOSINE"

(ix) FEATURE:

- (A) NAME/KEY: misc_difference
- (B) LOCATION: replace(20, "")
- (D) OTHER INFORMATION: /note= "W REPRESENTS EQUAL MIXTURE OF A AND T"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

AGGTNNANCT NCTCGAGTCW GG

22

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 38 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

CTATTAAC TAACTA CTAACCGTAA CAGTGGTGCC TTGCCCCA

38

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

AGGCTTACTA GTACAATCCC TGGGCACAAT

30

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

CCAGTTCCGA GCTCGTTGTG ACTCAGGAAT CT

32

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

CCAGTTCCGA GCTCGTGTG ACCGAGCCGC CC

32

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

CCAGTTCCGA GCTCGTGCTC ACCCAGTCTC CA

32

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

CCAGTTCCGA GCTCCAGATG ACCCAGTCTC CA

32

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

CCAGATGTGA GCTCGTGATG ACCCAGACTC CA

32

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

CCAGATGTGA GCTCGTCATG ACCCAGTCTC CA

32

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

CCAGTTCCGA GCTCGTGATG ACACAGTCTC CA

32

(2) INFORMATION FOR SEQ ID NO:25:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

GCAGCATTCT AGAGTTTCAG CTCCAGCTTG CC

32

(2) INFORMATION FOR SEQ ID NO:26:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 34 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

GGCCCGTCTA GAATTAACAC TCATTCCTGT TGAA

34

(2) INFORMATION FOR SEQ ID NO:27:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 37 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

GATCCTAGGC TGAAGGGAT GACCCTGCTA AGGCTGC

37

(2) INFORMATION FOR SEQ ID NO:28:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 35 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

ATTCAATAGT TTACAGGCAA GTGCTACTGA GTACA

35

(2) INFORMATION FOR SEQ ID NO:29:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 35 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

TTGGCTACGC TTGGGCTATG GTAGTAGTTA TAGTT

35

(2) INFORMATION FOR SEQ ID NO:30:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 35 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

GGTGCTACCA TAGGGATTAA ATTATTCAAA AAGTT

35

(2) INFORMATION FOR SEQ ID NO:31:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 18 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

TACGAGCAAG GCTTCTTA

18

(2) INFORMATION FOR SEQ ID NO:32:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 39 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

AGCTTAAGAA GCCTTGCTCG TAAACTTTTT GAATAATTT

39

(2) INFORMATION FOR SEQ ID NO:33:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 36 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

69

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

AATCCCTATG GTAGCACCAA CTATAACTAC TACCAT

36

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 35 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

AGCCCCAAGCG TAGCCAATGT ACTCAGTAGC ACTTG

35

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 34 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

CCTGTAAACT ATTGAATGCA GCCTTAGCAG GGTC

34

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 16 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

ATGGCCTTCA GCCTAG

16

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

CATTGGCTTAG A

21

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

70

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

TAGCATTAAC GTCCAATA

18

(2) INFORMATION FOR SEQ ID NO:39:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 26 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

ATATATTTTA GTAAGCTTCA TCTTCT

26

(2) INFORMATION FOR SEQ ID NO:40:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 23 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

GACAAAGAAC CGGTGAAAAC TTT

23

(2) INFORMATION FOR SEQ ID NO:41:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 35 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

GGGGGCCTCT TCGCTATTGC TTAAGAAGCC TTGCT

35

(2) INFORMATION FOR SEQ ID NO:42:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 43 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

AAACGACGGG CAGTGCCAAG TGACGGTGT GAAATTGTTA TCC

43

(2) INFORMATION FOR SEQ ID NO:43:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 43 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

GGCGAAAGGG AATTCTGCAA GGCGATTAAG CTTGGGTAAC GCC

43

(2) INFORMATION FOR SEQ ID NO:44:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 36 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

GGCGTTACCC AAGCTTTGTA CATGGAGAAA ATAAAG

36

(2) INFORMATION FOR SEQ ID NO:45:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

TGAAACAAAG CACTATTGCA CTGGCACTCT TACCGTTACC GT

42

(2) INFORMATION FOR SEQ ID NO:46:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

TACTGTTTAC CCCTGTGACA AAAGCCGCC AGGTCCAGCT GC

42

(2) INFORMATION FOR SEQ ID NO:47:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 44 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

72

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

TCGAGTCAGG CCTATTGTGC CCAGGGATTG TACTAGTGGA TCCG

44

(2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 38 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

TGGCGAAAGG GAATTCCGGAT CCACTAGTAC AATCCCTG

38

(2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 42 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

GGCACAAATAG GCCTGACTCG ACCAGCTGGA CCAGGGGGGC TT

42

(2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 42 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

TTGTCACAGG CGTAAACAGT AACGGTAACG GTAAGTGTGC CA

42

(2) INFORMATION FOR SEQ ID NO:51:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 42 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

GTGCAATAGT GCTTTGTTTC ACTTTATTTC CTCCATGTAC AA

42

(2) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

TAACGGTAAG AGTGCCAGTG C

21

(2) INFORMATION FOR SEQ ID NO:53:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 32 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

CACCTTCATG AATTCCGCAA GGAGACAGTC AT

32

(2) INFORMATION FOR SEQ ID NO:54:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 22 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

AATTCCGCAA GGAGACAGTC AT

22

(2) INFORMATION FOR SEQ ID NO:55:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 39 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

AATGAAATACTTATGCCCTA CGGGAGCCGC TGGATTGTT

39

(2) INFORMATION FOR SEQ ID NO:56:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 39 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

ATTACTCGCT GCCCAACCAG CCATGGCCGA GCTCGTGAT

39

74

(2) INFORMATION FOR SEQ ID NO:57:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 39 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

GACCCAGACT CCAGATATCC AACAGGAATG AGTGTAAAT

39

(2) INFORMATION FOR SEQ ID NO:58:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 13 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

TCTAGAACCG GTC

13

(2) INFORMATION FOR SEQ ID NO:59:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 45 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

TTCAGGTTGA AGCTTACCGCG TTCTAGAATT AACACTCATT CCTGT

45

(2) INFORMATION FOR SEQ ID NO:60:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 39 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

TGGATATCTG GAGTCTGGGT CATCACGAGC TCGGCCATG

39

(2) INFORMATION FOR SEQ ID NO:61:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 39 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

75

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

GCTGGTTGGG CACCGAGTAA TAACAATCCA GCGGCTGCC

39

(2) INFORMATION FOR SEQ ID NO:62:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 37 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

GTAGGCAATA GGTATTTCAT TATGACTGTC CTTGGCG

37

(2) INFORMATION FOR SEQ ID NO:63:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

TGACTGTCTC CTTGGCGTGT GAAATTGTTA

30

(2) INFORMATION FOR SEQ ID NO:64:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 36 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

TAACACTCAT TCCGGATGGA ATTCTGGAGT CTGGGT

36

(2) INFORMATION FOR SEQ ID NO:65:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

GCCAGTGCCA AGTGACCGGT TCTA

24

(2) INFORMATION FOR SEQ ID NO:66:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

ATATATTTTA GTAAGCTTCA TCTTCT

26

(2) INFORMATION FOR SEQ ID NO:67:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 23 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

GACAAAGAAC GCGTGAAAAC TTT

23

(2) INFORMATION FOR SEQ ID NO:68:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 76 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

CTGAACCTGT CTGGGACCAC AGTTGATGCT ATAGGATCAG ATCTAGAATT CATTAGAGA

60

CTGGCCTGGC TTCTGCG

76

(2) INFORMATION FOR SEQ ID NO:69:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 80 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:

TCGACCGTTG GTAGGAATAA TGCAATTAAT GGAGTAGCTC TAAATTAGA ATTCACTAC

60

ACCCAGTGCA TCCAGTAGCT

80

(2) INFORMATION FOR SEQ ID NO:70:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 27 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:70:

GGTAAACAGT AACGGTAAGA GTGCCAG

27

77

(2) INFORMATION FOR SEQ ID NO:71:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 54 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:71:

CGCCCTTCAGC CTAAGAACCG TAGTCCGGAA CGTCGTACGG GTAGGATCCA CTAG

54

(2) INFORMATION FOR SEQ ID NO:72:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 41 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:72:

CACCGGTTCG GGGAAATTAGT CTTGACCAGG CAGCCCAGGG C

41

(2) INFORMATION FOR SEQ ID NO:73:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 51 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:73:

ATTCCACACA TTATACGAGC CGGAAGCATA AAGTGTCAAG CCTGGGGTGC C

51

(2) INFORMATION FOR SEQ ID NO:74:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:74:

CTGCTCATCA GATGGCGGGA AGAGCTCGGC CATGGCTGGT TG

42

(2) INFORMATION FOR SEQ ID NO:75:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

78

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:75:

GAACAGAGTG ACCGAGGGGG CGAGCTCGGC CATGGCTGGT TG

42

I Claim:

1. A composition of matter comprising a plurality of cells containing diverse combinations of first and second DNA sequences encoding first and second polypeptides which form heteromeric receptors, one or both 5 of said polypeptides being expressed as fusion proteins on the surface of a cell.
2. The composition of claim 1, wherein said plurality of cells are E. coli.
3. The composition of claim 1, wherein said heteromeric receptors selected from the group consisting of antibodies, T cell receptors, integrins, hormone receptors and transmitter receptors.
4. The composition of claim 1, wherein said first and second DNA sequences encode functional portions of heteromeric receptors.
5. The composition of claim 4, wherein said first and second DNA sequences encode functional portions of the variable heavy and variable light chains of an antibody.
6. The composition of claim 1, wherein said cell produces filamentous bacteriophage.
7. The composition of claim 6, wherein said filamentous bacteriophage are selected from the group consisting of M13, fd and f1.
8. The composition of claim 6, wherein at least one of the encoded first or second polypeptides is expressed as a fusion protein with gene VIII.

9. A kit for the preparation of vectors useful for the coexpression of two or more DNA sequences encoding polypeptides which form heteromeric receptors comprising two vectors, a first vector having two pairs of restriction sites symmetrically oriented about a cloning site which can be combined with a second vector, having two pairs of restriction sites symmetrically oriented about a cloning site and in an identical orientation to that of the first vector, wherein one or both vectors contains sequences necessary for expression of polypeptides encoded by DNA sequences inserted in said cloning sites.

10. The kit of claim 9, wherein said first and second vectors are circular.

11. The kit of claim 9, wherein said expression peptides is as fusion proteins on the surface of a cell.

12. The kit of claim 9, wherein said cell produces filamentous bacteriophage.

13. The kit of claim 9, wherein said filamentous bacteriophage is selected from the group consisting of M13, fd and f1.

14. The kit of claim 13, wherein at least one of the DNA sequences is expressed as a fusion protein with gene VIII.

15. The kit of claim 9, wherein said two pairs of restriction sites are Hind III-Mlu I and Hind III-Mlu I.

16. A cloning system for the coexpression of two or more DNA sequences encoding polypeptides which form a heteromeric receptor, comprising a set of first vectors having a diverse population of first DNA sequences and a 5 set of second vectors having a diverse population second DNA sequences, said first and second vectors having two pairs of restriction sites symmetrically oriented about a cloning site for containing said first and second populations of DNA sequences so as to allow only the 10 operational combination of vector sequences containing said first and second DNA sequences.

17. The cloning system of claim 16, wherein said first and second vectors are circular.

18. The cloning system of claim 16, wherein said heteromeric receptors selected from the group consisting of antibodies, T cell receptors, integrins, hormone receptors and transmitter receptors.

19. The cloning system of claim 16, wherein said first and second DNA sequences encode functional portions of heteromeric receptors.

20. The cloning system of claim 19, wherein said first and second DNA sequences encode functional portions of the variable heavy and variable light chains of an antibody.

21. The cloning system of claim 16, wherein said coexpression of two or more DNA sequences encoding polypeptides which form a heteromeric receptor is on the surface of cell.

22. The cloning system of claim 16, wherein said cell produces a filamentous bacteriophage.

23. The cloning system of claim 22 wherein said filamentous bacteriophage selected from the group consisting of M13, fd and f1.

24. The cloning system of claim 23, wherein at least one of the DNA sequences is expressed as a fusion protein with the protein product of gene VIII.

25. The cloning system of claim 16, wherein said two pairs of restriction sites are Hind III-Mlu I and Hind III-Mlu I.

26. A plurality of expression vectors containing a plurality of possible first and second DNA sequences encoding polypeptides which form a heteromeric receptor exhibiting binding activity toward a preselected molecule,
5 said DNA sequence encoding heteromeric receptors being operatively linked to genes encoding surface proteins of a cell.

27. The expression vectors of claim 26, wherein said expression vectors are circular.

28. The expression vectors of claim 23, wherein said heteromeric receptors are selected from the group consisting of antibodies, T cell receptors, integrins, hormone receptors and transmitter receptors.

29. The expression vectors of claim 26, wherein said first and second DNA sequences encode functional portions of heteromeric receptors.

30. The expression vectors of claim 29, wherein said first and second DNA sequences encode functional portions of the variable heavy and variable light chains of an antibody.

31. The expression vectors of claim 26, wherein said cells produce filamentous bacteriophage.

32. The expression vectors of claim 26, wherein said filamentous bacteriophage are selected from the group consisting of M13, fd and f1.

33. The expression vectors of claim 32, wherein at least one of the encoded first or second polypeptides is expressed as a fusion protein with gene VIII.

34. A method of constructing a diverse population of vectors capable of expressing a diverse population of heteromeric receptors, comprising:

- 5 (a) operationally linking to a first vector a first population of diverse DNA sequences encoding a diverse population of first polypeptides, said first vector having two pairs of restriction sites symmetrically oriented about a cloning site;
- 10 (b) operationally linking to a second vector a second population of diverse DNA sequences encoding a diverse population of second polypeptides, said second vector having two pairs of restriction sites symmetrically oriented about a cloning site in an identical orientation to that of the first vector; and
- 15 (c) combining the vector products of step (a) and (b) under conditions which allow only the operational combination of vector sequences containing said first and second DNA sequences.
- 20

35. The method of claim 34, wherein said first and second vectors are circular.

36. The method of claim 34, wherein said heteromeric receptors are selected from the group consisting of antibodies, T cell receptors, integrins, hormone receptors and transmitter receptors.

37. The method of claim 34, wherein said first and second DNA sequences encode functional portions of the variable heavy and variable light chains of an antibody.

38. The method of claim 34, wherein said expression of a diverse population of heteromeric receptors is on the surface of a cell.

39. The method of claim 37, wherein said cell produces a bacteriophage.

40. The method of claim 39, wherein said filamentous bacteriophage is selected from the group consisting of M13, fd and fl.

41. The method of claim 34, wherein at least one of said first or second DNA sequences is expressed as a gene VIII fusion protein.

42. The method of claim 34, wherein said two pairs of restriction sites are Hind III-Mlu I and Hind III-Mlu I.

43. The method of claim 34, wherein said combining step further comprises:

5

(C1) restricting said first vector with a restriction enzyme recognizing one of the restriction sites encoded in said two pairs of restriction sites;

10

(C2) restricting said second vector with a different restriction enzyme recognizing the second restriction site encoded in said two pairs of restriction sites;

(C3) digesting the 3' ends of said restricted first and second vectors with an exonuclease; and

15

(C4) annealing said first and second vectors.

44. A method for selecting a heteromeric receptor exhibiting binding activity toward a preselected molecule from a population of diverse heteromeric receptors, comprising:

- 5 (a) operationally linking to a first vector a first population of diverse DNA sequences encoding a diverse population of first polypeptides, said first vector having two pairs of restriction sites symmetrically oriented about a cloning site;
- 10 (b) operationally linking to a second vector a second population of diverse DNA sequences encoding a diverse population of second polypeptides, said second vector having two pairs of restriction sites symmetrically oriented about a cloning site in an identical orientation to that of the first vector;
- 15 (c) combining the vector products of step (a) and (b) under conditions which allow only the operational combination of vector sequences containing said first and second DNA sequences.
- 20 (d) introducing said population of combined vectors into a compatible host under conditions sufficient for expressing said population of first and second DNA sequences; and
- 25 (e) determining the heteromeric receptors which bind to said preselected molecule.

45. The method of claim 44, wherein said first and second vectors are circular.

46. The method of claim 44, wherein said heteromeric receptors are selected from the group consisting of antibodies, T cell receptors, integrins, hormone receptors and transmitter receptors.

47. The method of claim 44, wherein said first and second DNA sequences encode functional portions of heteromeric receptors.

48. The method of claim 47, wherein said first and second DNA sequences encode functional portions of the variable heavy and variable light chains of an antibody.

49. The method of claim 44, wherein said expression of a diverse population of heteromeric receptors is on the surface of a cell.

50. The method of claim 49, wherein said cell produces a filamentous bacteriophage.

51. The method of claim 50, wherein said filamentous bacteriophage is selected from the group consisting of M13, fd and f1.

52. The method of claim 51, wherein at least one of said first or second DNA sequences is expressed as a gene VIII fusion protein.

53. The method of claim 44, wherein said two pairs of restriction sites are Hind III-Mlu I and Hind III-Mlu I.

54. The method of claim 44, wherein said combining step further comprises:

5

10

15

- (C1) restricting said first vector with a restriction enzyme recognizing one of the restriction sites encoded in said two pairs of restriction sites;
- (C2) restricting said second vector with a different restriction enzyme recognizing the second restriction site encoded in said two pairs of restriction sites;
- (C3) digesting the 3' ends of said restricted first and second vectors with an exonuclease; and
- (C4) annealing said first and second vectors.

55. A method for determining the nucleic acid sequences encoding a heteromeric receptor exhibiting binding activity toward a preselected molecule from a diverse population of heteromeric receptors, comprising:

- 5 (a) operationally linking to a first vector
 a first population of diverse DNA
 sequences encoding a diverse population
 of first polypeptides, said first
 vector having two pairs of restriction
 sites symmetrically oriented about a
 cloning site;
- 10 (b) operationally linking to a second
 vector a second population of diverse
 DNA sequences encoding a diverse
 population of second polypeptides, said
 second vector having two pairs of
 restriction sites symmetrically
 oriented about a cloning site in an
 identical orientation to that of the
 first vector;
- 15 (c) combining the vector products of step
 (a) and (b) under conditions which
 allow only the operational combination
 of vector sequences containing said
 first and second DNA sequences.
- 20 (d) introducing said population of combined
 vectors into a compatible host under
 conditions sufficient for expressing
 said population of first and second DNA
 sequences;
- 25
- 30

- (e) determining the heteromeric receptors which bind to said preselected molecule;
- 5 (f) isolating the nucleic acid sequences encoding said first and second polypeptides; and
- (g) sequencing said nucleic acid sequences.

56. The method of claim 55, wherein said first and second vectors are circular.

57. The method of claim 55, wherein said first heteromeric receptors selected from the group consisting of antibodies, T cell receptors, integrins, hormone receptors and transmitter receptors.

58. The method of claim 55, wherein said first and second DNA sequences encode functional portions of heteromeric receptors.

59. The method of claim 58, wherein said first and second DNA sequences encode functional portions of the variable heavy and variable light chains of an antibody.

60. The method of claim 55, wherein said expression of a diverse population of heteromeric receptors is on the surface of a cell filamentous bacteriophage selected from the group consisting of M13, fd and f1 and at 5 least one of said first or second DNA sequences is expressed as a gene VIII fusion protein.

61. The method of claim 55, wherein said cell produces filamentous bacteriophage.

91

62. The method of claim 61, wherein said filamentous bacteriophage is selected from the group consisting of M13, fd and fl.

63. The method of claim 62, wherein at least one of said first or second DNA sequences is expressed as a gene VIII fusion protein.

64. The method of claim 50, wherein said two pairs of restriction sites are Hind III-Mlu I and Hind III-Mlu I.

65. The method of claim 50, wherein said combining step further comprises:

(C1) restricting said first vector with a restriction enzyme recognizing one of the restriction sites encoded in said two pairs of restriction sites;

5

(C2) restricting said second vector with a different restriction enzyme recognizing the second restriction site encoded in said two pairs of restriction sites;

10

(C3) digesting the 3' ends of said restricted first and second vectors with an exonuclease; and

15

(C4) annealing said first and second vectors.

66. A vector comprising two copies of a gene encoding a filamentous bacteriophage coat protein, one copy of said gene capable of being operationally linked to a DNA sequence encoding a polypeptide of a heteromeric receptor
5 wherein said DNA sequence can be expressed as a fusion protein on the surface of said filamentous bacteriophage or as a soluble polypeptide.

67. The vector of claim 66, wherein said two copies of said gene encode substantially the same amino acid sequence but have different nucleotide sequences.

68. The vector of claim 66, wherein said one copy of said gene is expressed on the surface of said filamentous bacteriophage.

69. The vector of claim 66, wherein said bacteriophage coat protein is M13 gene VIII.

70. The vector of claim 66, wherein said vector has substantially the same sequence as that shown in Figure 2 (SEQ ID NO: 1).

71. A vector comprising sequences necessary for the coexpression of two or more inserted DNA sequences encoding polypeptides which form heteromeric receptors and two copies of a gene encoding a filamentous bacteriophage
5 coat protein, one copy of said gene capable of being operationally linked to one of said two or more inserted DNA sequences wherein said DNA sequence can be expressed as a fusion protein on the surface of said filamentous bacteriophage or as a soluble polypeptide.

72. The vector of claim 71, wherein said two copies of said gene encode substantially the same amino acid sequence but have different nucleotide sequences.

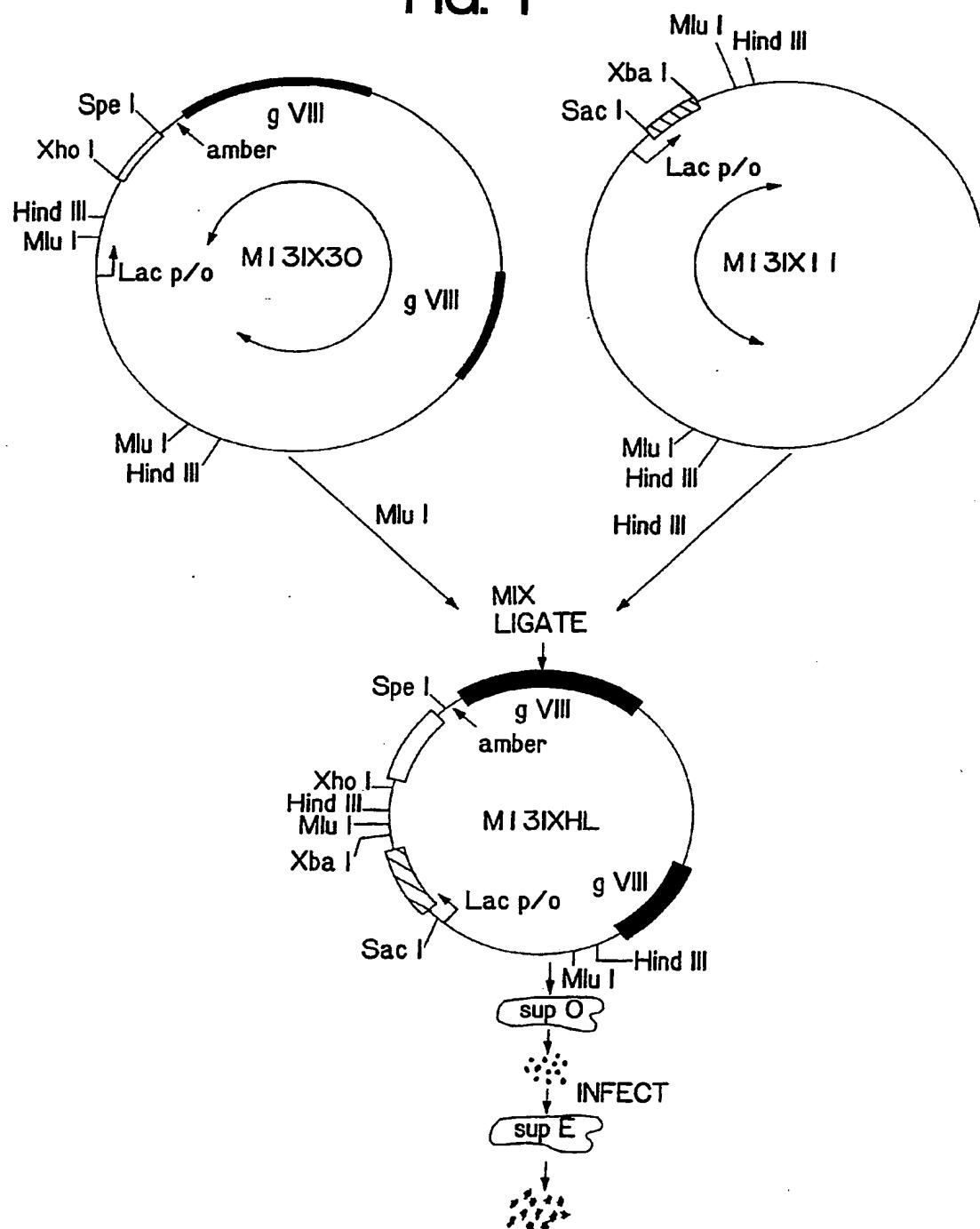
73. The vector of claim 71, wherein said one copy of said gene is expressed on the surface of said filamentous bacteriophage.

74. The vector of claim 71, wherein said bacteriophage coat protein is M13 gene VIII.

75. The vector of claim 71, wherein said vector has substantially the same sequence as that shown in Figure 6 (SEQ ID NO: 5).

1 / 11

FIG. 1



SUBSTITUTE SHEET

2 / 11

	10	20	30	40	50	60	
1	AATGCTACTA	CTATTAGTAG	AATTGATGCC	ACCTTTTCA	CTCGCGCCCC	AAATGAAAAT	60
61	ATAGCTAAC	AGGTTATTGA	CCATTTCGCA	AATGTATCTA	ATGGTCAAAC	TAAATCTACT	120
121	CGTTCGAGA	ATTGGGAATC	AACTGTTACA	TGGAATGAAA	CTTCAGACCA	CCGTACTTTA	180
181	GTTGCATATT	TAAAACATGT	TGAGCTACAG	CACCAAGATTC	AGCCAATTAAAG	CTCTAAAGCCA	240
241	TCTGCAAAAA	TGACCTCTTA	TCAAAAGGAG	CAATTAAAGG	TACTCTCTAA	TCCTGACCTG	300
301	TTGGAGTTG	CTTCCGGTCT	GGTTCGCTT	GAAGCTCGAA	TTAAAACGCG	ATATTGAAAG	360
361	TCTTCGGGC	TTCCCTCTTAA	TCTTTTGTAT	GCAATCCGCT	TTGCTTCTGA	CTATAATAGT	420
421	CAGGGTAAAG	ACCTGATTTT	TGATTTATGG	TCATCTCTGT	TTTCTGAACT	GTTTAAAGCA	480
481	TTTGAGGGGG	ATTCAATGAA	TATTTATGAC	GATTCCGAG	TATTGGACGC	TATCCAGTCT	540
541	AAACATTTTA	CTATTACCCC	CTCTGGCAAA	ACTCTTTTG	CAAAGGCCTC	TCGCTATTTT	600
601	GGTTTTTATC	GTCGTCGGT	AAACGAGGGT	TATGATAGTG	TTGCTCTTAC	TATGCCCTCGT	660
661	AATTCCCTTT	GGCGTTATGT	ATCTGCATTA	GTTGAATGTG	GTATTCCCTAA	ATCTCAACTG	720
721	ATGAATCTTT	CTACCTGTAA	TAATGTTGTT	CCGTTAGTTC	GTTTTATTAAG	CGTAGATTTT	780
781	TCTTCCCAAC	GTCCTGACTG	GTATAATGAG	CCAGTTCTTA	AAATGCATCA	AGGTAATTCA	840
841	CAATGATTA	AGTTGAAATT	AAACCATTC	AGCAGCTTTG	TTACGTTGAT	TTGGGTAATG	900
901	CTCGTCAGGG	CAAGCCTTAT	TCACTGAATG	AGCAGCTTTG	TTACGTTGAT	TTGGGTAATG	960
961	AATATCCGGT	TCTGTCAAG	ATTACTCTTG	ATGAAGGTCA	GCCAGCCAT	GCGCCCTGGTC	1020
1021	TGTACACCCT	TCATCTGTCC	TCTTCACAAAG	TTGGTCAGTG	CGGATCCCTT	ATGATTGACC	1080
1081	GTCTCGGCC	CGTTCGGGCT	AAGTAACATG	GAGCAGGTG	CGGATTCGCA	CACAAATTAT	1140
1141	CAGCGCATGA	TACAAATCTC	CGTTGTTACT	TGTTTCGCG	TTGGTTATAAT	CGCTGGGGGT	1200
1201	CAAAGATGAG	TGTTTTAGTG	TATTCCTTCG	CCTCTTCTG	TTTGGTTGG	TGCCTCTCGT	1260
1261	GTGGCATTAC	GTATTTTACC	CGTTTAATGG	AAACCTTCCTC	ATGAAAAAAAGT	CTTTAGTCCT	1320
1321	CAAAGCCTCT	GTAGCCGTTG	CTACCTCTGT	TCCGATGCTG	TCTTTCTCGT	CTGAGGGTGA	1380
1381	CGATCCCGCA	AAAGCGGCCCT	TTAACCTCCT	GCAAGCCTCA	GGCACCGAAT	ATATCGGTTA	1440
1441	TGCGTGGCG	ATGGTTGTTG	TCATTGTCGG	CGCACTATC	GGTATCAAGC	TGTTTAAGAA	1500
1501	ATTCACTCTG	AAAGCAAGCT	GATAAACCGA	TACAATTAAA	GGCTCTTTT	GGAGCCTTTT	1560
1561	TTTTTGAGA	TTTCAACACGT	AAAAAAATTA	TTATTCGCAA	TTCCCTTTAGT	TGTTCCCTTC	1620
1621	TATTCTCACT	CCGCTGAAAC	TGTTGAAGT	TGTTAGCAA	AAACCCATAC	AGAAAATTCA	1680
1681	TTTACTAACG	TCTGGAAAGA	CGACAAAAT	TTAGATCGT	ACGCTAACTA	TGAGGGTTGT	1740
1741	CTGTGGAATG	CTACAGGCGT	TGTTAGTTGT	ACTGGTACG	AAACTCAGTG	TTACGGTACA	1800
1801	TGGGTTCTA	TTGGGCTTGC	TATCCCTGAA	AATGAGGGTG	GTGGCTCTGA	GGGTGGCGGT	1860
1861	TCTGAGGGTG	GCGGTTCTGA	GGGTGGCGGT	ACTAAACCTC	CTGAGTACGG	TGATACACCT	1920
1921	ATTCGGGGT	ATACCTTATAT	CAACCTCTC	GACGCACTT	ATCCGCTCTG	TACTGAGCAA	1980
1981	AAACCCGCTA	ATCTCAATCC	TTCTTCTGAG	GAGTCCTCAG	CTCTTAATAC	TTTCATGTTT	2040
2041	CAGAATAATA	GGTTCGAAAA	TAGGCAGGGG	GCATTAACTG	TTTATACGGG	CACTGTTACT	2100
2101	CAAGGCCTG	ACCCCGTTAA	AACTTATTAC	CGATACACTC	CTGTATCATC	AAAAGCCATG	2160
2161	TATGACGCTT	ACTGGAACGG	TAAATTCTAGA	GAECTGCGCTT	TCCATTCTGG	CTTTATGAA	2220
2221	GATCCATTG	TTTGTGAATA	TCAAGGCAA	TCGTCCTGACC	TGCCCTCAACC	TCTGTCAAT	2280
2281	GCTGGCGCG	GCTCTGGTGG	TTGGTTCTGT	GGCGGCTCTG	AGGGTGGTGG	CTCTGAGGGT	2340
2341	GGCGGTTCTG	AGGGTGGCGG	CTCTGAGGGG	GGCGGTTCCG	GTGGTGGCTC	TGGTTCCGGT	2400
2401	GATTTTGATT	ATGAAAAGAT	GGCAACCGCT	AATAAGGGGG	CTATGACCGA	AAATGCCGAT	2460
2461	AAAACCGC	TACAGTCTGA	CGCTAAAGGC	AAACTTGATT	CTGTCGCTAC	TGATTACGGT	2520
2521	GCTGCTATCG	ATGGTTTCTAT	TGGTACGTT	TACCGGCTCTG	CTAATGGTAA	TGGTGCCTACT	2580
2581	GGTATTCTG	CTCGCTCTAA	TTCCCAAATG	GCTCAAGTCG	GTGACGGTGA	TAATTACACCT	2640
2641	TTAATGAATA	ATTTCGTC	ATATTTACCT	TCCCTCCCTC	AATCGGTTGA	ATGTGCCCT	2700
2701	TTTGTCTTA	GCCTGGTAA	ACCATATGAA	TTTCTTATTG	ATTGTGACAA	AATAAACTTA	2760
2761	TTCCGTTGGT	TCTTGTGTT	TCTTTTATAT	GTTGCCACCT	TTATGTATGT	ATTTTCTACG	2820
2821	TTTGCTAAC	TACTGCTGAA	TAAGGAGTCT	TAATCATGCC	AGTTCTTTG	GGTATTCCTG	2880
2881	TATTATTGCG	TTTCTCTCGGT	TTCTCTCTGG	TAACCTTGT	GGGCTATCTG	CTTACTTTTC	2940
2941	TTAAAAAGGG	CTTCGGTAAAG	ATAGCTATTG	CTATTTCATT	TTTCTTGT	CTTATTATTG	3000
3001	GGCTTAACTC	AATTCTTGTG	GGTTATCTCT	CTGATATTAG	CGCTCAATT	CCCTCTGACT	3060
3061	TTGTTCTAGGG	TGTTCACTT	ATTCTCCCTG	CTAATGCGCT	TCCCTGT	TATGTTATTG	3120
3121	TCTCTGTA	AAAGCTGTT	TTCTATTCTG	ACGTTAACAA	AAAAATCGTT	TCTTATTGTTG	3180
3181	ATTGGGATAA	ATATAATGGC	TGTTTATTGTT	GTAACTGGCA	ATTAGGCTC	TGAAAGAGACG	3240
3241	CTCGTTAGCG	TTGGTAAAGAT	TCAGGATAAA	ATTGTAGCTG	GGTGC	AAAGACTAAT	3300
3301	CTTGATTAA	GGCTTCAAAA	CCTCCCGCAA	GTGCGGAGGT	TCGCTAAAC	GCCTCGCGT	3360
3361	CTTGAATAC	CGGATAAGCC	TTCTATATCT	GATTGTTCTG	CTATTGGGCG	CGGTATATGAT	3420
3421	TCCTACGATG	AAAATAAAAAA	CGGCTTGCCT	GTTCTGATG	AGTGC	GGTAC	3480
3481	ACCCGTTCTT	GGAATGATAA	GGAAAGACAG	CCGATTATTG	ATTGGTTCT	ACATGCTCGT	3540
3541	AAATTAGGAT	GGGATATTAT	TTTCTTGT	CAGGACTTAT	CTATTGTTGA	TAAACAGGCG	3600
3601	CGTTCTGCAT	TAGCTGAACA	TGTTGTTTAT	TGTCGTCGTC	TGGACAGAA	TACTTTACCT	3660
3661	TTTGTGGT	CTTTATATTC	TCTTATTACT	GGCTCGAAA	TGCCCTCTGCC	TAATATTACAT	3720
3721	GTTGGCGTTG	TTAAATATGG	CGATTCTCAA	TTAAGCCCTA	CTGTTGAGCG	TTGGCTTTAT	3780

FIG. 2-1

SUBSTITUTE SHEET

3 / 11

3781	ACTGGTAAGA	ATTGTATAA	CGCATATGAT	ACTAACAGG	CTTTTCTAG	TAATTATGAT	3840
3841	TCCGGTGT	TTT ATTCTTATT	AACGCCATT	TTATCACAGC	GTCGGTATT	CAAACCAATT	3900
3901	AATTTAGGTC	AGAAGATGAA	GCTTACTAAA	ATATATTGAA	AAAAGTTTC	ACCGGTTCTT	3960
3961	TGTCTTGC	GA TTGGATTTC	ATCAGCATT	ACATATAGTT	ATATAACCCA	ACCTAAGCCG	4020
4021	GAGGTAAAAA	AGGTAGTCTC	TCAGACCTAT	GATTITGATA	AATTCACTAT	IGACTCTTCT	4080
4081	CAGCGCTT	TA ATCTAAGCTA	TCGCTATGTT	TCAGAAGGATT	CTAAGGGAAA	ATTAATTAAAT	4140
4141	AGCGACGATT	TACAGAAGCA	AGGTTATTC	CTCACATATA	TTGATTTATG	TACTGTTTCC	4200
4201	ATTAAAAAAG	GTAATTCAAA	TGAAATTGTT	AAATGTAATT	AATTITGTTT	TCTTGATGTT	4260
4261	TGTTTCATCA	TCTCTTTTG	CTCAGGTAAT	TGAAATGAAT	AATTTCGCCTC	TGCGCGATTT	4320
4321	TGTAACCTGG	TATTCAAAGC	AATCAGGCGA	ATCCGTTATT	GTTTCTCCG	ATGAAAAGG	4380
4381	TACTGTACT	GTATATTCTAT	CTGACGTTA	ACCTGAAAT	CTACGCAATT	TCTTATTTC	4440
4441	TGTTTACGT	GCTAATAATT	TTGATATGGT	TGGTCTCAATT	CCTTCATTA	TTCAGAAGTA	4500
4501	TAATCCAAAC	AATCAGGATT	ATATTGATGA	ATTGCCATCA	TCTGATAATC	AGGAATATGA	4560
4561	TGATAATTCC	GCTCCTTCTG	GTGGTTCTT	TGTTCCGCAA	AATGATAATG	TTACTCAAAAC	4620
4621	TTTTAAATT	ATAACGTT	GGGCAAAGGA	TTTAATACGA	GTTGTCAT	TGTTTGTAAA	4680
4681	GTCTAACT	TCTAAATTCT	CAAATGATT	ATCTATTGAC	GGCTCTAATC	TATTAGTTGT	4740
4741	TAGTGCACCT	AAAGATATT	TAGATAACCT	CTCTCAATT	CTTCTACTG	TTGATTTGCC	4800
4801	AACTGACCAG	ATATTGATTG	AGGGTTGAT	ATTGAGGTT	CAGCAAGGTG	ATGCTTAA	4860
4861	TTTTCATTT	GCTGCTGGCT	CTCAGCGTGG	CACTGTTGCA	GGCGGTGTTA	ATACTGACCG	4920
4921	CCTCACCTCT	GTTTTATCTT	CTGCTGGGG	TTCGTTCGGT	AATTTTAATG	GCGATGTTT	4980
4981	AGGGATCTAC	GTTCGCGCAT	TAAGAGCTAA	TAGCCTATTCA	AAAAATTGTT	CTGTGCCACG	5040
5041	TATTCTTACG	CTTTCAGGTC	AGAAGGGTT	TATCTCTGTT	GGCCAGAAATG	TCCCTTTAT	5100
5101	TACTGGTCGT	GTGACTGGTG	AATCTGCAA	TGAAATAAT	CCATTTCAGA	CGATTGAGCG	5160
5161	TCAAAATGTA	GGTATTTCCA	TGAGCTTTT	TCCTGTTGCA	ATGGCTGGCG	GTAATTATG	5220
5221	TCTGGATATT	ACCAAGCAAGG	CCGATAGTTT	GAGTTCTTCT	ACTCAGGCAA	GTGATGTTAT	5280
5281	TACTAATCAA	AGAAGTATTG	CTAACACGGT	TAATTGCGT	ATGGACAGCA	CTCTTTTACT	5340
5341	CGGTGGCCCTC	ACTGATTATA	AAAACACTTC	TCAAGATTCT	GGCGTACCGT	TCCTGTCTAA	5400
5401	AATCCCTTAA	ATCGGCCTCC	TGTTTAGCTC	CCGCTCTGAT	TCCAACGAGG	AAAGCACGTT	5460
5461	ATACGTGCTC	GTCAAAGCAA	CCATAGTAGC	CGGCCCTGTAG	CGGCGCATT	AGCGCGGCCG	5520
5521	GTGTGGTGGT	TACGCGCAGC	GTGACCTGTA	CACTTGCCAG	CGCCCTAGCG	CCCAGCTCTT	5580
5581	TCGCTTCTCT	CCCCCTTCTT	CTCGCACG	TCGCGGGCTT	CTCCCGTC	GCTCTAAATC	5640
5641	GGGGGCTCCC	TTTAGGGTT	CGATTTAGTG	CTTACGGCA	CCTCGACCCC	AAAAAACTTG	5700
5701	ATTTGGGTGA	TGGTTACGT	AGTGGGCAT	CGCCCTGATA	GACGGTTTTT	CGCCCCTTGA	5760
5761	CGTTGGAGTC	CACGTTCTTT	AATAGTGAC	TCTTGTTCCA	AACTGGAACA	ACACTCAACC	5820
5821	CTATCTGGG	CTAATTCTTT	GATTATAAG	GGATTTGCCC	GATTTCGGAA	CCACCATCAA	5880
5881	ACAGGATTTT	CGCCTGCTGG	GGCAAACACG	CYTGGACCGC	TTGCTGCAAC	TCTCTCAGGG	5940
5941	CCAGGCGGTG	AAAGGGCAATC	AGCTGTTGCC	CTCTCGCTG	GTAAGAAAGAA	AAACCAACCT	6000
6001	GGCGCCCAAT	ACGCAAACCG	CCTCTCCCG	CGCGTTGGCC	GATTCAATTAA	TGCAGCTGGC	6060
6061	ACGACAGGTT	TCCCGACTGG	AAAGCGGGCA	GTGAGCGCAA	CGCAATTAA	GTGAGTTAGC	6120
6121	TCACTCATTA	GGCACCCCGAG	GCTTACACT	TTATGCTTCC	GGCTCGTATG	TTGTGTTGGAA	6180
6181	TTGTGAGCGG	ATAACAAATT	CACACGCTC	ACTTGGCACT	GGCGCTGCTT	TTACAACGTC	6240
6241	GTGACTGGG	AAACCCCTGGC	GTTACCAAG	CTTTGTACAT	GGAGAAAATA	AAGTGAACAC	6300
6301	AAGCACTATT	GCACCTGGC	TCITACCGTT	ACC GTTACTG	TTTACCCCTG	TGACAAAAGC	6360
6361	CGCCCAGGTC	CAGCTGCTCG	AGTCAGGCT	ATTGTCGCCA	GGGGATTGTA	CTAGTGGATC	6420
6421	CTAGGCTGAA	GGCGATGACC	CTGCTAAGGC	TGCATTCAAT	AGTTTACAGG	CAAGTGTCTAC	6480
6481	TGAGTACATT	GGCTACGCTT	GGGCTATGGT	AGTAGTATA	TTGGGTGCTA	CCATAGGGAT	6540
6541	TAAATTATTTC	AAAAAGTTTA	CGAGCAAGGC	TTCTTAAGCA	ATAGCGAAGA	GGCCCGCACC	6600
6601	GATCGCCCTT	CCCCAACAGTT	GCGCAGCCTG	AATGGCAGAT	GGCGCTTTGC	CTGGTTTCCG	6660
6661	GCACCAAGAAG	CGGTGCCGG	AAGCTGGCTG	GAGTGCAGTC	TTCTCTGAGGC	CGATAACGGTC	6720
6721	GTCGCTCCCT	CAAACCTGGC	GATGCACGGT	TACGATGCCG	CCATCTACAC	CAACGTAACC	6780
6781	TATCCCATT	CGGTCAATT	GGCGTTGTT	CCCACGGAGA	ATCCGACGGG	TTGTTACTCG	6840
6841	CTCACATT	ATGTTGATGA	AAGCTGGCTA	CAGGAAGGCC	AGACGCGAAT	TATTTTTGAT	6900
6901	GGCGTCCCTA	TTGGTTAAAAA	AATGAGCTGA	TTTACACAAA	ATTTAACGCG	AATTTTAACA	6960
6961	AAATATTAAAC	GTTTACAATT	TAATATTG	CTTATACAAAT	CTTCTGTTT	TTGGGGCTTT	7020
7021	TCTGATTATC	AACCGGGGTA	CATATGTTG	ACATGCTAGT	TTTACGATTA	CCGTTCATCG	7080
7081	ATTCTCTTGT	TTGTCCTCAGA	CTCTCAGGCA	ATGACCTGAT	AGCCTTTGTA	GATCTCTCAA	7140
7141	AAATAGCTAC	CCTCTCCGGC	ATTAATTTAT	CAGCTAGAAC	GGTTGAATAT	CATATTGATG	7200
7201	GTGATTGAC	TGTCTCCGGC	CTTTCICACC	CTTTGAATC	TTTACCTACA	CATTACTCAG	7260
7261	GCATTGCATT	AAAAATATAT	GAGGGTTCTA	AAAATTTTTA	TCCTTGCGTT	GAAATAAAGG	7320
7321	CTTCTCCCGC	AAAAGTATT	CAGGGTCATA	ATGTTTTGG	TACAACCGAT	TTAGCTTAT	7380
7381	GCTCTGAGGC	TTTATTGCTT	AATTTGCTA	ATTCTTGC	TTGCGCTGAT	GATTATTG	7440
7441	ACGTT						7445

| 10 | 20 | 30 | 40 | 50 | 60

FIG. 2-2

4 / 11

	I	10	I	20	I	30	I	40	I	50	I	60
1	AATGCTACTA	CTATTAGTAG	AATTGATGCC	ACCTTTCA	CTCGCGCCCC	AAATGAAAAT	60					
61	ATAGCTAAC	AGGTATTG	CCATTGCGA	AATGTATCTA	ATGGTCAAAC	TAAATCTACT	120					
121	CGTCGAGA	ATTGGGAATC	AACTGTTACA	TGGAATGAAA	CTTCAGAGCA	CCGTACTTTA	180					
181	GTTGCATATT	AAAACATGT	TGAGCTACAG	CACCAAGATT	AGCAATTAAAG	CTCTAAGCCA	240					
241	TCCGCAAAAA	TGACCTCTTA	TCAAAAGGGAG	CAATTAAAGG	TACTCTCTAA	TCCTGACCTG	300					
301	TTGGAGTTTG	CTTCGGTCT	GGTTGCGCTT	GAAGCTCGAA	TTAAAACGCG	ATATTGAAAG	360					
361	TCTTTCGGGC	TTCTCTTAA	TCTTTTGAT	GCAATCCGCT	TTGCTCTCTGA	CTATAATAGT	420					
421	CAGGGTAAAG	ACCTGATTT	TGATTTATGG	TCATTCTCGT	TTTCTGAACT	GTTTAAAGCA	480					
481	TTTGGAGGGGG	ATTCAATGAA	TATTATGAA	TATTTATGAG	TATGGACGCG	TATCCAGTCT	540					
541	AAACATTAA	CTATTACCCC	CTCTGGCAA	ACTTCTTTG	CAAAAGCCTC	TCGCTATTTT	600					
601	GGTTTTATC	GTGCTCTGGT	AAACGAGGGT	TATGATAGTG	TTGCTCTTAC	TATGCTCGT	660					
661	AATTCCCTT	GGCGTTATGT	ATCTGCATTA	GTGGAATGTG	GTATTCCTAA	ATCTCAACTG	720					
721	ATGAATCTT	CTACCTGTAA	TAATGTTGTT	CCGTTAGTTC	AAATTTATTA	CGTAGATTTT	780					
781	TCTTCCCAC	GTCTGACTG	GTATAATGAG	CCAGTCTTAA	AAATCGCAT	AGGTAATTCA	840					
841	CAATGATTA	AGTTGAAATT	AAACCATCTC	AAGCCAATT	TACTACTCGT	TCTGGTGT	900					
901	CTCGTCAGGG	CAAGCCTTAT	TCACTGAATG	AGCAGCTTTG	TTACGTTGAT	TTGGGTAATG	960					
961	AATATCCGGT	TCTTGTCAAG	ATTACTCTT	ATGAAGGTC	GCCAGCCTAT	GCGCCTGGTC	1020					
1021	TGTACACCGT	TCATCTGTCC	TCTTCAAAAG	TTGGTCAGTT	CGGTTTCCCCT	ATGATTGACC	1080					
1081	GTCTGCGCCT	CGTTTCCGGCT	AAGTAACATG	GAGCAGGTCG	CGGATTTCGAA	CACAAATTAT	1140					
1141	CAGGGCGATGA	TACAAATCTC	CGTTGTACTT	TGTTTCGCGC	TTGGTATAAT	CGCTGGGGGT	1200					
1201	CAAAGATGAG	TGTTTTAGTG	TATTCCTTCG	CCTCTTTCGT	TTTAGGTTGG	TGCCCTTCGTA	1260					
1261	GTGGCATTAC	GTATTTTAC	CGTTTAATGG	AAACTCCCTC	ATGAAAAAAGT	CTTTAGTCCT	1320					
1321	CAAAGCCTCT	GTAGCCGTTG	CTACCCCTGT	TCCGATGCTG	TCTTCTCGT	CTGAGGGGTGA	1380					
1381	CGATCCCCGA	AAAGCGGCCA	TTAACCTCCC	GCAAGCCTCA	GCGACCGAAT	ATATCGGTTA	1440					
1441	TGCGTGGCG	ATGGTTGTTG	TCATTGTCGG	CGCAACTATC	GGTATCAAGC	TGTTTAAGAA	1500					
1501	ATTCACCTCG	AAAGCAAGCT	GATAAAACCGA	TACAATTAAA	GGCTCCCTTTT	GGAGCCTTTT	1560					
1561	TTTTTGGAGA	TTTTCACGT	GAAAAAAATTA	TTATTGCAA	TTCCCTTATAGT	TGTTCCCTTTC	1620					
1621	TATTCTCACT	CCCGTGAACAC	TGTTGAAAGT	TGTTTAGCAA	AACCCCATAC	AGAAAATTCA	1680					
1681	TATTACTAACG	TCTGGAAAGA	CGACAAACACT	TTAGATCGTT	ACGCTAACTA	TGAGGGTTGT	1740					
1741	CTGTGGATG	CTACAGGCGT	TGTTAGTTGT	ACTGGTGACG	AAACTCAGTG	TTACGGTACA	1800					
1801	GGGGTTCTTA	TTGGGCTTGC	TATCCCTGAA	AATGAGGGGTG	GTGGCTCTGA	GGGTGGCGGT	1860					
1861	TCTGAGGGTG	GGCGTTCTGA	GGGTGCGGT	ACTAAACCTC	TGAGTACCGG	TGATACACCT	1920					
1921	ATTCCGGGCT	ATACTTATAT	CAACCCCTCTC	GACGGCACTT	ATCCGCTCTG	TACTGAGCAA	1980					
1981	AAACCCGCTA	ATCTTAAATCC	TTCTTCTGAG	GAGTCTCAGC	CTCTTAAATAC	TTTCATGTTT	2040					
2041	CAGAAATAATA	GGTTCGAAA	TAGGCAGGGG	GCATTAACATG	TTTATACGGG	CACTGTTACT	2100					
2101	CAAGGCACTG	ACCCGTTAA	AACTTATTAC	CACTACACTC	CTGTATCATC	AAAAGCCATG	2160					
2161	TATGACGCTT	ACTGGAAACGG	TAAATTCTAGA	GAECTCGCCT	TCCATTCTGG	CTTTAATGAA	2220					
2221	GATCCATTCTG	TTTGTAATA	TCAAGGCCAA	TCGTCTGACC	TGCCCTCAACC	TCCTGTCAAT	2280					
2281	GCTGGCGCG	GCTCTGGTGG	TGGTTCTGTT	GGCGGCTCTG	AGGGTGGTGG	CTCTGAGGGT	2340					
2341	GGCGGTTCTG	AGGGTGGCGG	CTCTGAGGGG	GGCGGTTCCG	GTGGTGGCTC	TGTTCCGGT	2400					
2401	GATTTTGATT	ATGAAAAGAT	GGCAAACGCT	AATAAGGGGG	CTATGACCGA	AAATGCCGAT	2460					
2461	GAAAACCGC	TACGTCGTA	CGCTAAAGGC	AAACTGTGATT	CTGTCGCTAC	TGATTACGGT	2520					
2521	GCTGCTATCG	ATGGTTTCTAT	TGGTGACGTT	TCCGGCTTGC	CTAATGGTAA	TGGTGCTACT	2580					
2581	GGTATTGTT	CTGGCTCTAA	TTCCCAAAATG	GCTCAAGTCG	GTGACGGTGA	TAATTACACT	2640					
2641	TTAATGAATA	ATTCCTGTC	ATATTTACCT	TCCCTCCCTC	AATCGGTTGA	ATGTCGCCCT	2700					
2701	TTTGTCTTA	GCGCTGGTAA	ACCATATGAA	TTTCTTATG	TTGTGACAA	AATAAACTTA	2760					
2761	TTCCGTGGT	TCTTGTGCTT	TCTTGTATAT	GTTGCAACCT	TTATGTATGT	ATTTTCATG	2820					
2821	TTTGCTAAC	TACTCGTCAA	TAAGGAGTCT	TAATCATGCC	AGTCTTTTGT	GGTATTCCGT	2880					
2881	TATTATTGCG	TTTCCTCGGT	TTCTTCTGG	TAACTTTGTT	CGGCTATCTG	CTTACTTTTC	2940					
2941	TTAAAAAGGG	CTTCGGTAAAG	ATAGCTATTG	CTATTTCATT	GTTCCTTGC	CTTATTATTG	3000					
3001	GGCTTAAC	AATTCTTG	GGTTATCTCT	CTGATATTAG	CGCTCAATT	CCCTCTGACT	3060					
3061	TTGTTCAAGG	TGTTCAAGT	ATTCTCCCGT	CTAATGCGCT	TCCCTGT	TATGTTATTIC	3120					
3121	TCTCTGTAA	GGCTGCTATT	TTCTTGTATTT	ACGTTAAACAA	AAAAATCGTT	TCTTATTG	3180					
3181	ATTGGGATAAA	ATAATATGGC	TGTTTATTTT	GTAACTGGCA	AATTAGGCTC	TGGAAAGACG	3240					
3241	CTCGTTAGCG	TTGGTAAGAT	TCAGGATAAA	ATTGTAGCTG	GGTGCAAAAT	AGCAACTAAT	3300					
3301	CTTGATTAA	GGCTCAAAA	CCTCCCGCAA	GTCGGGAGGT	TCGCTAAAAC	GCCTCGCGTT	3360					
3361	CTTAGAATAC	CGGATAAGGC	TTCTATATCT	GATTGTCGTT	CTATTGGCG	CGGTAAATGAT	3420					
3421	TCCTACGATG	AAAAATAAAAA	CGGCTTGCCT	GTTCTCGATG	AGTGCCTG	TTGGTTAAAT	3480					
3481	ACCGTTCTT	GGAATGATAA	GGAAAGACAG	CCGATTATTG	ATGGTTTCT	ACATGCTCGT	3540					
3541	AAATTAGGAT	GGGATATTAT	TTTTCTTGT	CAGGACTTAT	CTATTGTTGA	TAAACAGGC	3600					
3601	CGTTCTGCA	TACGTGAAC	TGTTGTTTAT	TGTCGTCGTC	TGGACAGAAT	TACTTTACCT	3660					
3661	TTTGTGCGTA	CTTTATATT	TCTTATTACT	GGCTGCAAAA	TGCCTCTG	TAAATTACAT	3720					
3721	GGGGCGTTG	TTAAATATGG	CGATTCTCAA	TTAAGCCCTA	CTGTTGAGCG	TTGGCTTTAT	3780					
3781	ACTGGTAAGA	ATTGTATAA	CGCATATGAT	ACTAAACAGG	CTTTTTCTAG	TAATTATGAT	3840					

FIG. 3-1

SUBSTITUTE SHEET

5 / 11

3841	TCCGGTGT	TTCTTATT	AACGCCTAT	TTATCACACG	GTCGGTATT	CAAACCATT	3900
3901	AATTTAGGTC	AGAAGATGAA	GCTTACTAAA	ATATATTGAA	AAAAGTTTC	ACCGGTTCTT	3960
3961	TGTCTTGC	TTGGATTTC	ATCAGCATTT	ACATATAGTT	ATATAACCCA	ACCTAAGCCG	4020
4021	GAGGTAAAAA	AGTAGTCTC	TCAGACCTAT	GATTTGATA	AATTCACTAT	TGACTCTTCT	4080
4081	CAGCGCTTA	ATCTAACGTA	TCGCTATGTT	TTCAAGGATT	CTAAGGGAAA	ATTAATTAAT	4140
4141	AGCGACGATT	TACAGAACGA	AGGTATTCA	CTCACATATA	TTGATTTATG	TACTGTTTCC	4200
4201	ATTAATTAAG	GTAATTC	TGAAATTGTT	AAATGTAATT	AATTTGTTT	TCTTGATGTT	4260
4261	TGTTTACATCA	TCTTCTTTG	CTCAGGTAAT	TGAAATGAAT	AATTGCGCTC	TGCGCGATT	4320
4321	TGTAACCTGG	TATTC	AATCAGGCGA	ATCCGTTATT	GTTTCTCCG	ATGTAAGG	4380
4381	TACTGTTACT	GTATATTCA	CTGACGTTAA	ACCTGAAAAT	CTACGCAATT	TCTTTATTT	4440
4441	TGTTTACGT	GCTAATAATT	TTGATATGTT	TGGTCAATT	CTTCCTCATAA	TTCAGAAGTA	4500
4501	TAATCCAAAC	AATCAGGATT	ATATTGATGA	ATTGCCATCA	TCTGATAATC	AGGAATATGA	4560
4561	TGATAATTCC	GCTCTTCTG	GTGGTTCTT	TGTTCCGAA	AATGATAATG	TTACTCAAAC	4620
4621	TTTAAATT	AATAACGTT	GGGCAAAGGA	TTTAATACGA	GTTGTCGAAT	TGTTTGTAAA	4680
4681	GTCTAACT	TCTAAATCCT	CAAATGATT	ATCTATTGAC	GGCTCTAAATC	TATTAGTTGT	4740
4741	TAGTGCACCT	AAAGATATT	TAGATAACCT	TCCTCAATT	CTTTCTACTG	TTGATTTGCC	4800
4801	AACTGACCG	ATATTGATTG	AGGGTTGAT	ATTGAGGTT	CAGCAAGGTG	ATGCTTACA	4860
4861	TTTTCTATT	GCTGCTGGCT	CTCAGCGTGG	CACTGTTGCA	GGCGGTGTTA	ATACTGACCG	4920
4921	CCTCACCTCT	GT	CTGCTGGGG	TTGCTCGGT	ATTTTTAATG	GCGATGTTT	4980
4981	AGGGCTAC	GTTCGCGCAT	TAAGA	ACTGATT	TAGCATTCA	AAAATATTGT	5040
5041	TATCTTACG	CTTCAAGGTC	AGAAGGGTT	TATCTCTGTT	GGCCAGAAATG	TCCCCTTTAT	5100
5101	TACTGGTCG	GTGACTGGT	AATCTGCAA	TGAAATAAT	CCATTTCAGA	CGATTGAGCG	5160
5161	TCAAAATGTA	GGTATTTC	TGAGCGTTT	TCCTGTTGCA	ATGGCTGGCG	GTAATATTGT	5220
5221	TCTGGATATT	ACCAAGCAAGG	CCGATAGTTT	GAGTTCTCT	ACTCAGGCAA	GTGATGTTAT	5280
5281	TACTAATCAA	AGAAGTATTG	CTAACACGGT	TAATTGCGT	GATGGACAGA	CTCTTTACT	5340
5341	CGGTGGCCTC	ACTGATTATA	AAAACACTTC	TCAAGATT	GGCGTACCGT	TCCTGTCTAA	5400
5401	AATCCCTTA	ATCGGCTCC	TGTTAGCTC	CCGCTCTGAT	TCCAACGAGG	AAAGCACGTT	5460
5461	ATACGTGCTC	GTCAAAGCAA	CCATAGTAG	CGCCCTGTAG	GGCGCATT	AGCGCGGCGG	5520
5521	GTGTGGTGGT	TACGCGCAGC	GTGACCGCT	CACTGCGAG	CGCCCTAGCG	CCCCTCTCCT	5580
5581	TCGTTTCTT	CCCTCCTT	CTCGCCACGT	TCGCGGGCTT	TCCCCTGCAA	GCTCTAAATC	5640
5641	GGGGGCTCCC	TTT	AGTGGGTT	CGATTTAGTG	CTTACGGCA	CCTCGACCCC	5700
5701	ATTTGGGTGA	TGGTTACGT	AGTGGGCCAT	CGCCCTGATA	GACGGTTTTT	CGCCCTTTGA	5760
5761	CGTTGGAGTC	CACGTTCTT	AATAGTGGAC	TCTTGTCCA	AACTGGAACA	ACACTCAACC	5820
5821	CTATCTCGGG	CTATTCTT	GATTATAAG	GGATTGTTGCC	GATTGCGGAA	CCACCATCAA	5880
5881	ACAGGATT	CGCCTGCTGG	GGCAAACACAG	CGTGGACCGC	TTGCTGCAAC	TCTCTCAGGG	5940
5941	CCAGGCGGTG	AAAGGCAATC	AGCTGTTGCC	CGTCTCGT	GTGAAAAGAA	AAACCACCC	6000
6001	GGCGCCCAAT	ACGAAACCG	CCTCTCCCCG	CGCGTTGGCC	GATTCAATTAA	TGCAGCTGGC	6060
6061	ACGACAGGTT	TCCGACTGG	AAAGCGGCA	GTGAGCGCAA	CGCAATTAAAT	GTGAGTTAGC	6120
6121	TCACTCATT	GGCACCCAG	GCTTACAT	TTATGCTTCC	GGCTCGTATG	TTGTGTTGAA	6180
6181	TTGTGAGCG	ATAACATT	CACAGCCAA	GGAGACAGTC	ATAATGAAAT	ACCTATTGCC	6240
6241	TACGGCAGCC	GCTGGATTGT	TATTACTCGC	TGCCCAACCA	GCCATGGCCG	AGCTCGTGT	6300
6301	GACCCAGACT	CCAGATATCC	AACAGGAATG	AGTGTAAATT	CTAGAACGCG	TCACTTGGCA	6360
6361	CTGGCCGTG	TTTACAACG	TCGTGACTGG	GAAAACCTG	CGFTTACCC	AGCTTAATCG	6420
6421	CCTTGCAGAA	TTCCCTT	CCAGCTGGCG	TAATAGCGAA	GAGGGCCCGA	CCGATCGCCC	6480
6481	TTCCAACACG	TTGCGCAGCC	TGATTGGCGA	ATGGCECTT	GCCTGGTTTC	CGGCACCCAGA	6540
6541	AGCGGTGCCG	GAAAGCTGGC	TGGAGTGC	TCTTCTGAG	GCCGATACGG	TCGTGTC	6600
6601	CTCAAACG	CAGATGCACG	GTTACGATGC	GCCCACATC	ACCAACGTAA	CCTATCCC	6660
6661	TACGGTCAAT	CCGCCGTTG	TTCCCACGGA	GAATCCGACG	GGTTGTTACT	CGCTCACATT	6720
6721	TAATGTTGAT	GAAAGCTGGC	TACAGGAAGG	CCAGACGCGA	ATTATTTTG	ATGGCGTTCC	6780
6781	TATTGGTTAA	AAAATGAGCT	GATTTAACAA	AAATTTAACG	CGAATTTAA	AAAATATTA	6840
6841	ACGTTTACAA	TTTAAATATT	TGCTTACAC	ATCTTCTGT	TTTGGGGCT	TTTCTGATTA	6900
6901	TCAACGGGG	TACATATGAT	TGACATGCA	GTTTTACGAT	TACCGTTCAT	CGATTCTCTT	6960
6961	GTTTGCTCCA	GACTCTCAGG	CAATGACCTG	ATAGCCTT	AGATCTCT	AAAAAATAGCT	7020
7021	ACCCCTCTCG	GCATTAAATT	ATCAGCTAGA	ACGGTTGAAT	ATCATATTGA	TGGTATTGTT	7080
7081	ACTGTCTCG	GCCTTCTCA	CCCTTTGAA	TCTTACCTA	CACATTACTC	AGGCATTGCA	7140
7141	TTTAAATAT	ATGAGGGTTC	TTAAATTTT	TATCCTTGC	TTGAAATAAA	GGCTTCTCCC	7200
7201	GAAAAGTAT	TACAGGGTCA	TAATGTTT	GGTACAACCG	ATTAGCTTT	ATGCTCTGAG	7260
7261	GCTTATTG	TTAATT	TAATTCTT	CCTTGCGCTG	ATGATT	GGATGTT	7317

FIG. 3-2

SUBSTITUTE SHEET

6 / 11

	10	20	30	40	50	60
1	AATGCTACTA	CTATTAGTAG	AATTGATGCC	ACCTTTCAAG	CTCGCGCCCC	AAATGAAAAT
61	ATAGCTAAC	AGGTTATTGA	CCATTTCGGA	AATGTATCTA	ATGGTCAAAC	TAAATCTACT
121	CGTTCGCAGA	ATTGGGAATC	AACTGTACA	TGGAATGAAA	CTTCCAGACA	CCGTACTTTA
181	GTTGCATATT	AAAACATGT	TGAGCTACAG	CACCAAGATT	AGCAATTAAAG	CTCTAAGCCA
241	TCTGAAAAAA	TGACCTCTTA	TCAAAAGGAG	CAATTAAAGG	TACTCTCTAA	TCCTGACCTG
301	TTGGAGTTTG	CTTCCTCTAA	GGTTTCGCTT	GAAGCTCGAA	TTAAAACGCG	ATATTGAAAG
361	TCTTCGGGC	TTCCTCTAA	TCTTTCTGAT	GCAATCCGCT	TTGCTTCTGA	CTATAATAGT
421	CAGGGTAAAG	ACCTGATT	TGATTTATGG	TCATCTCGT	TTTCTGAACT	GTTAAAGCA
481	TTTGAGGGGG	ATTCATGAA	TATTTATGAC	GATTCGGCAG	TATGGACGCG	TATCCAGTCT
541	AAACATTTA	CTATTACCCC	CTCTGGCAA	ACTTCTTTG	CAAAAGCCTC	TCGCTATTT
601	GGTTTTTATC	GTCGCTGTG	AAACGAGGGT	TATGATAGTG	TTGCTCTTAC	TATGCTCGT
661	AATTCCCTT	GGCGTTATGT	ATCTGATTA	GTTGAATGTG	GTATTTCTAA	ATCTCAACTG
721	ATGAATCTT	CTACCTGTAA	TAATGTTGTT	CCGTTAGTTC	GTTTTATTA	CGTAGATTTT
781	TCTTCCCAAC	GTCCTGACTG	GTATAATGAG	CCAGTTCTTA	AAATCGCATA	AGGTAAATTCA
841	CAATGATTAA	AGTTGAAATT	AAACCACATC	AAGCCCAATT	TACTACTCGT	TCTGGTGT
901	CTCGTCAGGG	CAAGCCTT	TCACTGAATG	AGCAGCTTG	TTACGTTGAT	TTGGGTAATG
961	AATATCCGGT	TCTTGTCAAG	ATTACTCTG	ATGAAGGTG	GGCCAGCTAT	GCGCCTGGTC
1021	TGTACACCGT	TCATCTGTCC	TCTTCAAA	TTGGTCAGTT	CGGTTCCCTT	ATGATTGACC
1081	GTCTGCGCCT	CGTTCGGCT	AAGTAACATG	GAGCAGGTCG	CGGATTTCGA	CACAATTTAT
1141	CAGGCATG	TACAAATCTC	CGTTGTACTT	TGTTTCGCGC	TTGGTATAAT	CGCTGGGGGT
1201	CAAAGATGAG	TGTTTAGTG	TATTCTCTG	CCTCTTCG	TTTAGGTTGG	TGCCTTCGTA
1261	GTGGCATTAC	GTATTTTAC	CGTTTAATGG	AAACTCTCTC	ATGAAAAAGT	CTTATGTCCT
1321	CAAAGCCTT	GTAGCCGTG	TCACCCCTG	TCCGATGCTG	TCTTCGCTG	CTGAGGGTGA
1381	CGATCCCGCA	AAAGCGGCCT	TTAACCTCC	GCAAGCCTC	GCGACCGAAT	ATATCGGTTA
1441	TGCGTGGCG	ATGGTTGTTG	TCATTGTCGG	CGCAACTATC	GGTATCAAGC	TGTTTAAGAA
1501	ATTACACCTG	AAAGCAAGCT	GATAAACCGA	TACAATTAAA	GGCTCCTTTT	GGAGCCTTTT
1561	TTTTGGAGA	TTTCAACGT	GAAAAAAATTA	TTATTGCAA	TTCTCTTCTAGT	TGTTCCCTTC
1621	TATTCTCACT	CCGCTGAAAC	TGTTGAAAGT	TGTTTAGCAA	AACCCCATAC	AGAAAATTCA
1681	TTTACTAACG	TCTGGAAAGA	CGACAAAAC	TTAGATCGTT	ACGCTAACTA	TGAGGGTTGT
1741	CTGTGGAATG	CTACAGGCGT	TGTTAGTTGT	ACTGGTACG	AAACTCAGTG	TTACGGTACA
1801	TGGGTTCTA	TTGGGCTTG	TATCCCTGAA	AATGAGGGGTG	GTGGCTCTGA	GGGTGGCGGT
1861	TCTGAGGGTG	GCGGGCTCTGA	GGGTGGCGGT	ACTAACCTC	CTGAGTACGG	TGATACACCT
1921	ATTCCGGGCT	ATACTTATAT	CAACCCCTC	GACGGCACTT	ATCCGCGCTG	TACTGAGCAA
1981	AACCCCGCTA	ATCCTAATCC	TTCTCTTGTAG	GAGTCAGC	CTCTTAAATAC	TTTCATGTTT
2041	CAGAATAATA	GGTTCCGAAA	TAGGCAGGGG	GCATTAAC	TTTATACGGG	CACTGTTACT
2101	CAAGGCAC	ACCCCGTTAA	AACTTATTAC	CACTACACTC	CTGTATCATC	AAAAGCCATG
2161	TATGACGCTT	ACTGGAAAC	TAAATTCA	GACTGCGCTT	TCCATTCTGG	CTTTAATGAA
2221	GATCCATTG	TTTGTGAATA	TCAAGGCAA	TCGTCGAC	TGCCTCAAC	TCCTGTCAT
2281	GCTGGCGCG	GCTCTGGTGG	TGGTTCTGTT	GGCGGCTCTG	AGGGTGGTGG	CTCTGAGGGT
2341	GGCGGTTCTG	AGGGTGGCGG	CTCTGAGGGG	GGCGGTTCCG	GTGGTGGCTC	TGGTTCCGGT
2401	GATTTTGATT	ATGAAAAGAT	GGCAAACG	AATAAGGGG	CTATGACCGA	AAATGCGAT
2461	AAAAACGCGC	CTACGCTG	CGCTAAAGGC	AAACTTGATT	CTGTCGCTAC	TGATTACGGT
2521	GCTGCTATCG	ATGGTTTCAT	TGGTACGTT	TCCGGCCTTG	CTAATGGTAA	TGGTGCCTACT
2581	GGTGATT	CTGGCTCTAA	TTCCCAAATG	GCTCAAGTCG	GTGACGGTGA	TAATTCACCT
2641	TTAATGAATA	ATTICCGTCA	ATATTAC	TCCCTCC	AATCGGTTGA	ATGTCGCCCT
2701	TTTGTCTT	GCGCTGGTAA	ACCATATGAA	TTTTCTATTG	TTTGTGACAA	AATAAAACTTA
2761	TTCCTGCGTG	TCTTTCGTT	TCTTTTATAT	GTTGCGCACCT	TTATGTTATGT	ATTTTCTACG
2821	TTTGCTAAC	TACTGCGTAA	TAAGGAGTC	TAATCATGCC	AGTTCTTTTG	GGTATTCCGT
2881	TATTATTGCG	TTTCTCTCGGT	TTCCCTCTGG	TAACTTGTT	GCGTATCTG	CTTACTTTTC
2941	TTAAAAAGGG	CTTGGTAAAG	ATAGCTATG	CTATTCTT	GTTCCTCTG	CTTATTATTG
3001	GGCTTAAC	AATTCTTG	GGTTATCT	CTGATATTAG	CGCTCAATT	CCCTCTGACT
3061	TTGTTCA	TTGTCAGTTA	ATTCTCCCGT	CTAATGCG	TCCCTGTTT	TATGTTATT
3121	TCTCTGTA	AAAGCTATT	TTCAATT	ACGTTAAACA	AAAAATCGTT	TCTTATTG
3181	ATTGGGATAA	ATAATATGGC	TGTTTATT	GTAACTGGCA	AATTAGGCTC	TGGAAAGACG
3241	CTCGTAGCG	TTGGTAAAGAT	TCAGGATAAA	ATTGAGCTG	GGTGCAAAAT	AGCAACTAAT
3301	CTTGATT	GGCTTCAAAA	CCTCCCGCA	GTCGGGAGGT	TCGCTAAAC	GCCTCGCGTT
3361	CTTAGAATAC	CGGATAAGCC	TTCTATATCT	GATTGCTTG	CTATTGGCG	CGGTAATGAT
3421	TCCTACGAT	AAAATAAAAA	CGGCTTGCTT	GTTCTCGATG	AGTGCCTG	TTGGTTAAT
3481	ACCCGTTCT	GGAATGATAA	GGAAAGAGACG	CCGATTATTG	ATTTGTTTCT	ACATGCTCGT
3541	AAATTAGGAT	GGGATATTAT	TTTTCTT	CAGGACTT	CTATTGTTGA	TAAACAGGCG
3601	CGTTCTGCA	TAGCTGAAAC	TGTTGTTTAT	TGTCGCTG	TGGACAGAA	TACTTACCT
3661	TTTGTGCGTA	CTTTATATT	TCTTATTACT	GGCTGAAAA	TGCGCTCG	TAAATTACAT
3721	GTTGGCGTTG	TTAAATATGG	CGATTCTCAA	TTAAGCC	CTGTTGAGCG	TTGGCTTAT
3781	ACTGGTAAGA	ATTGGTATAA	CGCATATGAT	ACTAAACAGG	CTTTTCTAG	TAATTATGAT

FIG. 4-1
SUBSTITUTE SHEET

7 / 11

3841	TCCGGTGT	TTCTTATT	AACGCC	TTATCACACG	GTCGGTATT	CAAACC	3900
3901	AAATTAGG	CTGAAAGATGAA	GCTTACTA	AAATATTTGA	AAAAGTTT	ACGC	3960
3961	TGTCTTGCG	TTGGATTTC	ATCAGCA	ATCATATAGT	ATATAACCC	ACCTAAGCC	4020
4021	GAGGTTAAA	AGGTAGTCTC	TCAGAC	GATTTGATA	AATTCACTAT	TGACTCTT	4080
4081	CAGCGTCT	ATCTAACGCA	TCGCTATG	TTCAAGGATT	CTAAGGGAA	ATTAATTAA	4140
4141	AGCGACGATT	TACAGAACG	AGGTTATTCA	CTCACATATA	TTGATTATG	TACTGTTCC	4200
4201	ATTAAGGAA	GTAATTCAA	TGAAATTG	AAATGTAA	AATTGTTG	TCTTGATGTT	4260
4261	TGTTTCATCA	TCTCTTTG	CTCAGG	TGAAATGA	AATTGCGCTC	TGCGCGATT	4320
4321	TGTAACCTGG	TATTCAAAGC	AATCAGGCG	ATCCGTT	TTTCTCCC	ATGTAAGG	4380
4381	TACTGTTACT	GTATATTCA	CTGACGTTA	ACCTGAAA	CTACGCAATT	TCTTATTTC	4440
4441	TGTTTACGT	GCTAATAATT	TTGATATG	TGGTTCA	CTTCCATAA	TTCAGAAGTA	4500
4501	TAATCCAAC	AATCAGGATT	ATATGATG	ATTGCGCATC	TCTGATA	AGGAATATGA	4560
4561	TGATAATTCC	GCTCTTCTG	GTGGTTCT	TGTTCCGAA	AATGATAATG	TTACTCAAAC	4620
4621	TTTAAATTA	AATAACGTT	GGGCAAGGA	TTTAACGCA	TTGTCGA	TGTTGTAAA	4680
4681	GTCTAATAC	TCTAAATCCT	CAAATGTATT	ATCTATTGAC	GGCTCTAAC	TATTAGTTG	4740
4741	TAGTGCACCT	AAAGATATT	TAGATAAC	TCCTCAATT	CTTCTACTG	TTGATTGCC	4800
4801	AACTGACAG	ATATGATTG	AGGGTTG	ATTTGAGG	CAGCAGG	ATGCTTACA	4860
4861	TTTTCAATT	GCTGCTGG	CTCAGC	CACTGTTG	GGCGGTGTTA	ATACTGACCG	4920
4921	CCTCACCTT	TTTTTATCTT	CTGCTGGT	TTCTGTTG	ATTTTAATG	GCGATGTT	4980
4981	AGGGCTATCA	GTTCGCGC	TAAGACTAA	TAGCCATTCA	AAAATATTG	CTGTGCCACG	5040
5041	TATTCTTACG	CTTCAGGTC	AGAAGGGT	TATCTCTG	GGCCAGAATG	TCCCTTTAT	5100
5101	TAAGGTCG	GTGACTGGT	AAATCTG	TGTAATAAT	CCATTCTCAGA	CGATTGAGCG	5160
5161	TCAAAATGTA	GGTATTTCA	TGAGCG	TCTGTTG	ATGGCTGG	GTAATTATG	5220
5221	TCTGGATATT	ACCAGCAAGG	CCGATAGTT	GAGTTCTT	ACTCAGGCA	GTGATGTT	5280
5281	TACTAATCAA	AGAAGTATTG	CTAACACG	TAATTGCG	GATGGACAGA	CTCTTTACT	5340
5341	CGGTGGCCTC	ACTGATTATA	AAAACACTC	TCAAGATTCT	GGCGTACCG	TCCTGCTAA	5400
5401	ATACCCCTTA	ATCGCCCTC	TGTTTAGCTC	CCGCTCTG	TCCAACGAGG	AAAGCACGTT	5460
5461	ATACGTGCTC	GTCAAGCA	CCATAGTACG	CGGCCCTG	CGCGCATTAA	AGCGCGGC	5520
5521	TGTTGGTGT	TACGCGCAG	GTGACCGCTA	CACTGCCAG	CGCCCTAGCG	CCCCTCCTT	5580
5581	TCGCTTCTT	CCCTTCTT	CTCGCCACG	TCGCCGCTT	TCCCCGTCAA	GCTCTAAATC	5640
5641	GGGGGCTCC	TTTGGGTT	CGATTAGT	CTTACGGCA	CCTCGACCCC	AAAAAAACTG	5700
5701	ATTTGGG	TGGTTACGT	AGTGGG	CGCCCTG	ACGGGTTT	CGCCCTTGA	5760
5761	CGTTGGAGTC	CACGTTCTT	AAATAGTGG	TCTGTTCCA	AACTGGAAACA	ACACTCAAAC	5820
5821	CTATCTCGG	CTATTCTT	GATTATAAG	GGATTG	GATTTCGGAA	CCACCATCAA	5880
5881	ACAGGATT	CGCCTGCTG	GGCAAACCA	CGTGGACCG	TTGTCGAAC	TCTCTCAGGG	5940
5941	CCAGGCGGTG	AAGGCAATC	AGCTGTTG	CGTCTCG	GTGAAAAGAA	AAACCAACCT	6000
6001	GGCAGCCCA	ACGCAAGG	CCTCTCCCG	CGCCTGG	GATTCTAA	TGCAGCTGGC	6060
6061	ACGACAGG	TCCCAGT	AAAGCGG	GTGAGCG	CACCAATTAA	GTGAGTTAGC	6120
6121	TCACTCATT	GGCACCC	GCTTAC	TTATGTT	GGCTG	TTGTGTTGAA	6180
6181	TTGTGAGCGG	ATAACAATT	CACACG	ACTTGGCA	GGCGTCG	TTACAACGTC	6240
6241	GTGACTGGG	AAACCTGG	GTTACCA	CTTGTAC	GGAGAAAATA	AAGTGAACAA	6300
6301	AAGCACT	GCACG	TCTTACG	ACTGTTT	CGTCTGG	AAGCCCAGGT	6360
6361	CGAGCTG	GAGTGG	TCCCCCTG	ACCCCTC	AAAGCAC	CTGGGGGCAC	6420
6421	AGCAGGCC	GGCTG	TCAAGACTAA	TTCCCCG	CGGTGACGG	GTCGTGGAAC	6480
6481	TCAGGCGCC	TGACAGC	CGTGCACACC	TTCCCCG	TCTTACAGTC	CTCAGGACTC	6540
6541	TACTCCCTCA	GCAGCGT	GACCG	TCCAGCAG	TGGGCACCC	GACCTACATC	6600
6601	TGCAACG	ATCACAGG	CAGCAACAG	AAAGTGG	AGAAAGCAGA	GCCCCAAATCT	6660
6661	TGTACTAGT	GATCTACCC	GTACGACG	CCGGACTAC	CTTCTTAGG	TGAAGGGC	6720
6721	GACCTGCTA	AGGCTG	CAATAGT	CAGGCAAG	CTACTGAGT	CATTGGCTAC	6780
6781	GCTTGGGCTA	TGGTAGT	TATAGTTG	GCTACCATAG	GGATTAAATT	ATTCAAAAG	6840
6841	TTTACGAGCA	AGGCTT	AGCAATAGC	AAGAGGCCG	CCAGCGATCG	CCTTCCCAAC	6900
6901	AGTTGCGCAG	CCTGAATGG	GAATGGCG	TTGCC	TCCGGCACCA	GAAGCGGTG	6960
6961	CGGAAAGCTG	GCTGGAGT	GATCTTCTG	AGGCCG	GGTCGTC	CCCTAAACT	7020
7021	GGCAGATGCA	CGGTTACG	GCGCC	ACACCAACG	AACCTATCCC	ATTACGGTCA	7080
7081	ATCCGCCGTT	TGTTCCCACG	GAGAATCCG	CGGGTTG	CTCGCTCAC	TTAATGTTG	7140
7141	ATGAAAGCTG	GCTACAGG	GGCCAGG	GAATTAT	TGATGGCG	CCTATTGG	7200
7201	AAAAAATGAG	CTGATTAA	AAAATTA	CGCGAATT	AACAAAAT	TAACGTTAC	7260
7261	ATTTAAATA	TTTGT	CAATCTT	TTTTGGGG	CTTTCTG	TATCAACCG	7320
7321	GGTACATATG	ATTGACATG	TACTTT	ATTACG	ATCGATT	TTGTTGCTC	7380
7381	CAGACTCTCA	GGCAATGAC	TGATAGC	TGATAGT	TCAAAATAG	CTACCCCTC	7440
7441	CGGCATTAAT	TTATCAG	GAACGGG	ATATCAT	GATGGT	TGACTGTC	7500
7501	CGGCCTTCT	CACCTTTG	ATCTTAC	TACACATTAC	TCAGGCA	CATTAAAAT	7560
7561	ATATGAGGGT	TCTAAAATT	TTTACCTT	CGTTGAAATA	AAGGCTT	CCGCAAAAGT	7620
7621	ATTACAGGGT	CATAATGTTT	TTGGTACAAC	CGATTAGCT	TTATGCTC	AGGCTTATT	7680
7681	GCTTAATT	GCTAATTCTT	TGCTTGC	GTATGATT	TTGGACG		7729

| 10 | 20 | 30 | 40 | 50 | 60 |

FIG. 4-2
SUBSTITUTE SHEET

8 / 11

	10	20	30	40	50	60
1	AATGCTACTA	CTATTAGTAG	AATTGATGCC	ACCTTTCAAG	CTCGCGCCCC	AAATGAAAAT
61	ATAGCTAAC	AGGTTATTGA	CCATTTCGCA	AATGTATCTA	ATGGTCAAAC	TAAATCTACT
121	CGTTCGAGA	ATTGGGAATC	AACTGTTACA	TGGAATGAAA	CTTCAGACAGA	CCGTACTTTA
181	GTTGCATATT	TAACACATGT	TGAGCTACAG	CACCAAGATTC	AGCAATTAAAG	CTCTAACGCCA
241	TCGGCAAAAA	TGACCTCTTA	TCAAAAGGAG	CAATTAAGGAG	TACTCTCTAA	TCCTGACCTG
301	TTGGAGTTTG	CTTCCTTAA	GGTTCGCTT	GAAGCTCGAA	TTAAAACGCG	ATATTGAAAG
361	TCTTCGGGC	TTCCTCTAA	TCTTTTGTAT	GCAATCCGCT	TTGCTCTGA	CTATAATAGT
421	CAGGGTAAAG	ACCTGATTTT	TGATTTATGG	TCACTCTCGT	TTTCTGAAC	GTTAAAGCA
481	TTTGAGGGGG	ATTCAATGAA	TATTTATGAC	GATTCGGCAG	TATGGGACGC	TATCCAGTCT
541	AAACATTAA	CTATTACCCC	CTCTGGCAA	ACTTCTTTG	CAAAGGCTC	TCGCTATTAA
601	GGTTTTTATC	GTCTGCTGGT	AAACGAGGGT	TATGATAGTG	TTGCTCTTAC	TATGCTCGT
661	AATTCCCTT	GGCGTTATGT	ATCTGCATTA	GTTGAATGTG	GTATTCTAA	ATCTCAACTG
721	ATGAATCTT	CTACCTGTAA	TAATGTTGTT	CCGTTAGTTC	TTTTTATTA	CGTAGATTTT
781	TCTTCCCAC	GTCCTGACTG	GTATAATGAG	CCAGGTTCTTA	AAATCGCATA	AGGTAATTCA
841	CAATGATTA	AGTTGAAATT	AAACCATCTC	AAGCCAAAT	TACTACTCGT	TCTGGTGT
901	CTCGTCAGGG	CAAGCCTTAT	TCACTGAATG	AGCAGCTTTG	TTACGTTGAT	TTGGGTAATG
961	AATATCCGGT	TCTTGTCAAG	ATTACTCTTG	ATGAAGGTCA	GCCAGCCTAT	GCGCCTGGTC
1021	TGTACACCGT	TCATCTGTCC	TCTTCAAAAG	TTGGTCAGTT	CGGTTCCCTT	ATGATTGACC
1081	GTCTGCGCT	CGTTCCGGT	AAGTAACATG	GAGCAGGTGCG	CGGATTCGCA	CACAATTAT
1141	CAGGGCATGA	TACAAATCTC	CGTTGTACTT	TGTTTCGCGC	TTGGTATAAT	CGCTGGGGGT
1201	CAAAGATGAG	TGTTTGTG	TATTCCTTCG	CCTCTTCTGT	TTAGGTTGG	TGCCTTCGTA
1261	GTGGCATTAC	GTATTTTACC	CGTTTAATGG	AAACTTCCTC	ATGAAAAAAGT	CTTAACTCCT
1321	CAAAGCCTCT	GTAGCCGTTG	CTACCTCTGT	TCCGATGCTG	TCTTCGCTG	CTGAGGGTGA
1381	CGATCCCGCA	AAAGCCGCTC	TTAACCTCCCT	GCAAGCCTCA	GCGACCGAAT	ATATCGTTA
1441	TGCGTGGCG	ATGGTTGTT	TCATTGTCGG	CGCAACTATC	GGTATCAAGC	TGTTAAAGAA
1501	ATTCACCTCG	AAAGCAAGCT	GATAAAACCGA	TACAATTAAA	GGCTCCTTT	GGAGCCTTT
1561	TTTTGGAGA	TTTCAACGT	AAAAAAATTA	TTATTGCAA	TTCCCTTGT	TGTTCTTT
1621	TATTCTCACT	CCGCTGAAAC	TGTTGAAAGT	TGTTAGCAA	AACCCCATAC	AGAAAATTCA
1681	TTTACTAACG	TCTGGAAAGA	CGACAAAATG	TTAGATCGT	ACGCTAACTA	TGAGGGTTGT
1741	CTGTGGAATG	CTACAGGCCT	TGTTGTTGT	ACTGGTACG	AAACTCAGTG	TTACGGTACA
1801	TGGGTTCTA	TTGGGCTTGC	TATCCCTGAA	AATGAGGGTG	GTGGCTCTGA	GGGTGGCGGT
1861	TCTGAGGGTG	GCGGITCTGA	GGGTGGCGGT	ACTAACCTC	CTGAGTACGG	TGATACACCT
1921	ATTCCGGGCT	ATACCTTAT	CAACCCCTCTC	GACGCACTT	ATCCGCTGG	TACTGAGCAA
1981	AACCCCGCTA	ATCTTAATCC	TTCTCTGAG	GAGTCAGTC	CTCTTAATAC	TTTCATGTTT
2041	CAAGATAATA	GGTTCGAAA	TAGGCAGGGG	GCATTAAC	TTATACGGG	CACTGTTACT
2101	CAAGGCACTG	ACCCCGTTAA	AACTTATTAC	CAGTACACTC	CTGTATCATC	AAAAGCCATG
2161	TATGACGCTT	ACTGGAACGG	TAAATTGCAA	GAATGCGCTT	TCCATTCTGG	CTTTAAATGAA
2221	GATCCATTG	TTTGTAATA	TCAAGGCAA	TCGTCAGTC	TCGCTCAACC	TCCTGTCAAT
2281	GCTGGCGCG	GCTCTGGTGG	TGGTTCTGGT	GGCGCGCTG	AGGGTGGTGG	CTCTGAGGGT
2341	GGCGGTTCTG	AGGGTGGCGG	CTCTGAGGGG	GGCGGTTCCG	GTGGTGGCTC	TGGTTCCGGT
2401	GATTTTGATT	ATGAAAAGAT	GGCAGAACGCT	AATAAGGGGG	CTATGACCGA	AAATGCCGAT
2461	AAAAACGCGC	TACAGTCTGA	CGCTAAAGGC	AAACCTTGATT	CTGTCGCTAC	TGATTACGGT
2521	GCTGCTATCG	ATGGTTTCAT	TGGTACGTT	TCCGGCTCTG	CTAATGGTAA	TGGTGTACT
2581	GGTGGATTGG	CTGGCTCTAA	TTCCCTAAATG	GCTCAAGTC	GTGACGGTGA	TAATTACCT
2641	TTAATGAATA	ATTTCCGTCA	ATATTTACTC	TCCTCCCTC	AATCGGTTGA	ATGTCGCCCT
2701	TTTGCTTTA	CGCGTGGTAA	ACCATATGAA	TTTTCTATTG	ATTGTGACAA	AATAAAACTA
2761	TTCCGTGGT	TCTTTCGTT	TCTTTTATAT	GTTGCCACCT	TTATGATGTT	ATTTTCTACG
2821	TTTGCTAAC	TACTGCGTAA	TAAGGAGTCT	TAATCATGCC	AGTTCTTTG	GGTATTCCGT
2881	TATTATTGCG	TTTCCTCGGT	TTCTCTGTT	TAACCTTGT	CGGCTATCTG	CTTACTTTT
2941	TTAAAAAGGG	CTTCGGTAAAG	ATAGCTATTG	CCTGTTCTT	GCTCTTATTA	TTGGGCTTAA
3001	CTCAATTCTT	GTGGGTTATC	TCTCTGATAT	TAGCGCTCAA	TTACCCCTCTG	ACTTTGTTCA
3061	GGGTGTTCA	TTAATTCTCC	CGTCTAATGC	GCTTCCTCTG	TTTTATGTTA	TTCTCTCTGT
3121	AAAGGCTGT	ATTTCTATT	TTGACGTTAA	ACAAAAAAATC	TTTCTTATT	TGGATTGGGA
3181	TAAAATATA	GGCTGTTTAT	TTTGTAAC	GCAAAATTAGG	CTCTGGAAAG	ACGCTCGTTA
3241	GCGTTGGTAA	GATTCAAGGAT	AAAATTGAG	CTGGGTGCAA	AATAGCAACT	AATCTTGATT
3301	TAAGGCTTCA	AAACCTCCCG	CAAGTCGGGA	GGTCGCTAA	AACGCCCTCGC	GTTCCTTAGAA
3361	TACCGGATAA	GCCTTCTATA	TCTGATTGTC	TTGCTATTGG	GCGCGGTAAT	GATTCTACG
3421	ATGAAAATAA	AAACGGCTTG	CTTGTCTCG	ATGAGTGCAG	TACTTGGTTT	AATACCGTT
3481	CTTGGAAATG	TAAGGAAAGA	CAGCCGATTA	TTGATTGGTT	TCTACATGCT	CGTAAATTAG
3541	GATGGGATAT	TATTTTCTT	GTTCAAGGACT	TATCTATTG	TGATAAAACAG	GCGCGTTCTG
3601	CATTAGCTGA	ACATGTTGTT	TATTGTCGTC	GTCTGGACAG	AATTACTTTA	CCTTTTGTG
3661	GTACTTTATA	TTCTCTTATT	ACTGGCTGA	AAATGCCCTCT	GCCTAAATTA	CATGTTGGCG
3721	TTGTTAAATA	TGGCGATTCT	CAATTAAGCC	CTACTGTTGA	GCGTTGGCTT	TATACTGGTA
3781	AGAATTGTA	TAACCGCATAT	GATACTAAC	AGGCTTTTTC	TAGTAATTAT	GATTCCGGTG

FIG. 5-1
SUBSTITUTE SHEET

9 / 11

3841	TTTATTCTTA	TTAACGCCT	TATTTATCAC	ACGGTCGGTA	TTTCAAACCA	TTAAATTTAG	3900
3901	GTCAGAAGAT	GAAGCTTACT	AAAATATATT	TGAAAAAGTT	TTCACGCGTT	CTTTGTCTTG	3960
3961	CGATTGGATT	TGCATCAGCA	TTTACATATA	GTTATATAAC	CCAACCTAAG	CCGGAGGTTA	4020
4021	AAAAGGTAGT	CTCTCAGACC	TATGATTITG	ATAAATTCAAC	TATTGACTCT	TCTCAGCGTC	4080
4081	TTAATCTAAG	CTATCGCTAT	GTTTTCAAGG	ATTCTAAGGG	AAAATTAAATT	AATAGCGACG	4140
4141	ATTTACAGAA	GCAAGGTTAT	TCACTCACAT	ATATTGATTT	ATGTAATCTT	TCCATTAAAA	4200
4201	AAGGTAATT	AAATGAAATT	GTAAAATGTA	ATTAAATTITG	TTTCTTGTAT	GTTTGTTTCA	4260
4261	TCATCTTCTT	TTGCTCAGGT	AATTGAAATG	AATAATTCGC	CTCTGCGCGA	TTTTGTAACT	4320
4321	TGGTATTCAA	AGCAATCAGG	CGAATCGCTT	ATTGTTTCTC	CCGATGTAA	AGGTACTGTT	4380
4381	ACTGTATATT	CATCTGACGT	TAACACGTAA	AATCTACGCA	ATTTCTTAT	TTCTGTTTA	4440
4441	CGTGTAAATA	ATTTGATAT	GTTGGTTCA	ATTCTTCCA	TAATTCAAGAA	GTATAATCCZ	4500
4501	AACAATCAGG	ATTATATTGA	TGAATTGCCA	TCATCTGATA	ATCAGGAATA	TGATGATAAT	4560
4561	TCCGCTCCTT	CTGGTGGTTT	CTTTGTTCCG	CAAATGATA	ATGTTACTCT	AACTTTAAA	4620
4621	ATTAATAACG	TTCGGGCAAA	GGATTTATA	CGAGTTGTCG	AATTGTTTGT	AAAGTCTAAT	4680
4681	ACTTCTAAAT	CCTCAATGTT	ATTATCTATT	GACGGCTCTA	ATCTATTAGT	TGTTAGTGCA	4740
4741	CCTAAAGATA	TTTCAAGATAA	CCTTCCTCAA	TTCTTTCTA	CTGTTGATT	GCCAACGTGAC	4800
4801	CAGATAATTG	TTGAGGGTTT	GATATTGAG	GTTCAAGCAAG	GTGATGCTT	AGATTTTCA	4860
4861	TTTGTGCTG	GCTCTCAGCG	TGGCACTGTT	GCAGGCGGTG	TTAATACTGA	CCGCCCTCACC	4920
4921	TCTGTTTAT	CTCTGCTGG	TGGTTTCGTC	GGTATTTTA	ATGGCGATGT	TTTAGGGCTA	4980
4981	TCAGTTGCG	CATTAAGAC	TAATAGCCT	TCAAAATAT	TGTCGTGCC	ACGTATTCTT	5040
5041	ACGCTTTCAG	GTCAAGAAGGG	TTCTATCTCT	GTTGCCAGA	ATGTCCTTT	TATTACTGGT	5100
5101	CGTGTGACTG	GTGAATCTGC	CAATGTAAT	AATCATTTC	AGACGATTGA	GCGTCAAAAT	5160
5161	GTAGGTATT	CCATGAGCGT	TTTCCCTGTT	GCAATGGCTG	GCGGTAATAT	TGTTCTGGAT	5220
5221	ATTACCAAGCA	AGGCCGATAG	TTTGAGTTCT	TCTACTCAGG	CAAGTGTATG	TATTACTAAT	5280
5281	CAAAGAAGTA	TTGTCACAAAC	GGTTAATTG	CGTGTGGAC	AGACTCTTTT	ACTCGGTGGC	5340
5341	CTCACTGATT	ATAAACACAC	TTCTCAAGAT	TCTGGCGTAC	CGTTCCGTGTC	AAAAATCCCT	5400
5401	TTAATCGGCC	TCCTGTTAG	CTCCCGCTCT	GATTCACACG	AGGAAAGCAC	GTTATACGTG	5460
5461	CTCGTCAAAG	CAACCATACT	ACCGCCCTG	TAGCGGCGCA	TTAAGCGCGG	CGGGGTGTGGT	5520
5521	GGTTACCGC	AGCGTACCG	CTACACTG	CAGGCCCTA	GCGCCCGCTC	CTTTCGCTTT	5580
5581	TTTCGCTGC	TGGGGCAAAAC	CAGCGTGGAC	CGCTGCTGC	AACCTCTCA	GGGCCAGGCG	5940
5941	GTGAAGGGCA	ATCAGCTGTT	GCCCCGCTCG	CTGGTGAAAAA	AAAAAAACAC	CCTGGCGCCC	6000
6001	AATAACGAAA	CCGCCCTCTCC	CCGCGCGTTG	GCCGATTCAT	TAATGCAGCT	GGCACGACAG	6060
6061	GTTCCCGAC	TGGAAAGCGG	GCAGTGAAGCG	CAACGCAATT	AATGTGAGTT	AGCTCACTCA	6120
6121	TTAGGCACCC	CAGGCTTTAC	ACTTTATGCT	TGGCGCTCGT	ATGTTGTGTG	GAATTGTGAG	6180
6181	CGGATAACAA	TTTCACACGC	CAAGGAGACA	GTCATAATGA	AATAACCTATT	GCCTACGGCA	6240
6241	GCCGCTGGAT	TGTTTATTAC	CGCTGCCAA	CCAGCCATGG	CCGAGCTCTT	CCCGCCATCT	6300
6301	GATGAGCACT	TGAAATCTGG	AACTGCCTCT	GTTGTGTGCC	TGCTGAATAA	CTTCTATATCCC	6360
6361	AGAGAGGCCA	AAAGTACAGTG	GAAGGTTGAT	AACGCCCTCC	AATCGGGTAA	CTCCCAGGAG	6420
6421	AGTGTACACG	AGCAGGACAG	CAAGGACAGC	ACCTACAGCG	TCAGCAGCAC	CCTGACGCTG	6480
6481	AGCAAAGCAG	ACTACGAGAA	ACACAAAGTC	TACGCCGTGCG	AAAGTCACCCA	TCAGGGCCTG	6540
6541	AGCTGCCCG	TCACAAAGAG	CTTCAACAGG	GGAGAGTGT	CTAGAACGGG	TCACTTGGCA	6600
6601	CTGGCCGCTG	TTTACAACG	TCGTGACTGG	GAAAACCTG	GCCTTACCCA	AGCTTAATCG	6660
6661	CCTTCAGAA	TTCCCTTTCG	CCAGCTGGCG	TAATAGCGAA	GAGGCCCGCA	CCGATCGCCC	6720
6721	TTCCCAACAG	TTGCGCAGCC	TGAATGGCGA	ATGGCGCTTT	GCCTGGTTTC	CGGCACCCAGA	6780
6781	AGCGGTGCG	CAAAGCTGGC	TGGAGTGCAG	TCTTCTGAG	GCCGATACGG	TCGTGTCCCC	6840
6841	CTCAAACCTGG	CAGATGCACG	GTTACGATGC	GCCCCATCTAC	ACCAACGTA	CCTATCCCCAT	6900
6901	TACGGTCAAT	CCGCCGTTTG	TTCCCACGGA	GAATCCGACG	GGTGTGACT	CGCTCACATT	6960
6961	TAATGTTGAT	GAAAGCTGGC	TACAGGAAGG	CCAGACGCGA	ATTATTTTGT	ATGGCGTTCC	7020
7021	TATTGGTTAA	AAAATGAGCT	GATTTAACAA	AAATTTAAC	CGAATTTTAA	AAAATATTA	7080
7081	ACGTTTACAA	TTTAAATATT	TGCTTATACA	ATCTCTGT	TTTTGGGGCT	TTTCTGATTA	7140
7141	TCAACCGGGG	TACATATGAT	TGACATGCTA	GTTTTACGAT	TACCGTTCAT	CGATTCTCTT	7200
7201	GTGGTCTCCA	GACTCTCAGG	CAATGACCTG	ATAGCCCTTG	TAGATCTCTC	AAAAATAGCT	7260
7261	ACCCCTCTCG	GCATTAATT	ATCAGCTAGA	ACGGTTTGAAT	ATCATATTGA	TGGTGATTG	7320
7321	ACTGTCTCCG	GCCTTTCTCA	CCCTTTGAA	TCTTACCTA	CACATTACTC	AGGCATTGCA	7380
7381	TTTAAATAT	ATGAGGGTTC	AAAAAATTT	TATCCTTGCG	TTGAAATAAA	GGCTTCTCCC	7440
7441	GCAAAAGTAT	TACAGGGTCA	TAATGTTTT	GGTACAACCG	ATTTAGCTT	ATGCTCTGAG	7500
7501	GCTTTATTGC	TTAATTTTGC	TAATTCTTIG	CCTTGCGCTGT	ATGATTATT	GGATGTT	7557

FIG. 5-2

10 / 11

	1	10	20	30	40	50	60
1	AATGCTACTA	CTATTAGTAG	AATTGATGCC	ACCTTTCAAG	CTCGCGCCCC	AAATGAAAAT	60
61	ATAGCTAAC	AGGTTATTGA	CCATTGCGA	AATGTATCTA	ATGGTCAAAC	TAAATCTACT	120
121	CGTTCGCAGA	ATTGGGAATC	AACTGTTACA	TGGAATGAAA	CTTCCAGACA	CCGTAACCTTA	180
181	GTTGCATATT	TAAAACATGT	TGAGCTACAG	CACCAAGATT	AGCAATTAAAG	CTCTAAGCCA	240
241	TCTGCAAAAA	TGACCTCTTA	TCAAAAGGGAG	CAATTAAAGG	TACTCTCTAA	TCCTGACCTG	300
301	TTGGAGTTTG	CTTCGGCTCT	GGTTCGCTT	GAAGCTCGAA	TTAAAACGCG	ATATTTGAAG	360
361	TCTTCGGGC	TTCTCTTAA	TCTTTTGTAT	GCAATCCGCG	TTGCTTCTGA	CTATAATAGT	420
421	CAGGGTAAAG	ACCTGATTTT	TGATTTATGG	TCATTCTCGT	TTTCTGAAC	GTTTAAACGA	480
481	TTTGAGGGGG	ATTCAATGAA	TATTTATGAC	GATTCCGCG	TATTGGACGC	TATCCAGTCT	540
541	AAACATTTA	CTTATACCCC	CTCTGGCAA	ACTTCTTTG	CAAAAGCCTC	TCGCTATTT	600
601	GGTTTTTATC	GTGCGCTGGT	AAACGAGGGT	TATGATAGTG	GTGCTTCTAC	TATGCCCTCGT	660
661	AAATCCCTT	GGCGTTATGT	ATCTGCATTA	GTTGAATGTC	GTATTCTAA	ATCTCAACTG	720
721	ATGAATCTT	CTACCTGTAA	TAATGTTGTT	CCGTTAGTTC	GTTTATTAAT	CGTAGATT	780
781	TCTTCCCAAC	GTCCTGACTG	GTATAATGAG	CCAGTTCTTA	AAATCGCATA	AGGTAATTCA	840
841	CAATGATTA	AGTTGAAATT	AAACCATCTC	AAGGCCAATT	TACTACTCGT	TCTGGTGT	900
901	CTCGTCAGGG	CAAGCCTTAT	TCACTGAATG	AGCAGCTTTG	TTACGTTGAT	TTGGGTTAATG	960
961	AAATATCCGGT	TCTTGTCAAG	ATTACTCTG	ATGAAGGTC	GCCAGCCTAT	GCGCCTGGTC	1020
1021	TGTACACCCT	TCATCTGTCC	TCTTCAAG	TTGGTCAGTT	CGGTTCCCTT	ATGATTGACC	1080
1081	GTCTGCGCCT	CGTTCGGCT	AAGTAACATG	GAGCAGGTCG	CGGATTTCGA	CACAATTAT	1140
1141	CAGGCGATGA	TACAAATCTC	CGTTGACTT	TGTTTTCGCG	TGGTATAAT	CGCTGGGGGT	1200
1201	CAAAGATGAG	GTGTTTAGT	TATTTTGTCT	CCTCTTTCG	TTAGGTTGG	TGCCCTCGTA	1260
1261	GTGGCATTAC	GTATTCTTAC	CGTTTAATGG	AAACTCTCTC	ATGAAAAAGT	CTTCTAGT	1320
1321	CAAAGCCTCT	GTAGCCGTTG	CTACCCCTCGT	TCCGATGCTG	TCTTCGCTG	CTGAGGGTGA	1380
1381	CGATCCCGCA	AAAGCGGCCT	TTAACCTCCT	GCAAGCCTCA	GCGACCBAAT	ATATCGTTA	1440
1441	TGCGTGGCG	ATGGTTGTTG	TCATTGTCGG	CGCAACTATC	GGTATCAAGC	TGTTTAAGAA	1500
1501	ATTACCTCG	AAAGCAAGG	GATAAACCGA	TACAAATAAA	GGCTCCCTT	GGAGCCTTT	1560
1561	TTTGGGAGA	TTTCAACGT	AAAAAAATTA	TTATTCGCAA	TTCTCTTATG	TGTTCTTTC	1620
1621	TATTCTCACT	CCGCTGAAAC	TGTTGAAAGT	TGTTTAGCAA	AACCCCATAC	AGAAAATTCA	1680
1681	TTTACTAACG	TCTGAAAGA	CGACAAAAC	TTAGATCGTT	ACGCTAACTA	TGAGGGTTGT	1740
1741	CTGTGGAATG	CTACAGCGT	TGAGTTGT	ACTGTTGACG	AAACTCAGTG	TTACGGTACA	1800
1801	TGGGTTCTCA	TTGGGCTTGC	TATCCCTGAA	AAATGAGGGT	GTGGCTTCTGA	GGGTGGCGGT	1860
1861	TCTGAGGGTG	GCCTTCTGA	GGGGTGGCGGT	ACTAAACCTC	CTGAGTACGG	TGATACACCT	1920
1921	ATTCCGGGCT	ATACCTATAT	CAACCCCTCTC	GACGGCATT	ATCCGCCTGG	TACTGAGCAA	1980
1981	AAACCCCGCTA	ATCCTAATCC	TTCTCTTGTAG	GAGTCTCAGC	CTCTTAATAC	TTTCATGTTT	2040
2041	CAGAATAATA	GGTTGGGAA	TAGGCAGGGG	GCATTAACCTG	TCTTATACGGG	CACTGTTACT	2100
2101	CAAGGCACTG	ACCCCGTTAA	AACCTTATAC	CACTACACTC	CTGTTATCATC	AAAAGCCTATG	2160
2161	TATGACGCTT	ACTGGAACGG	TAAATTCAAGA	GACTGCGCTT	TCCTATTCTGG	CTTAAATGAA	2220
2221	GATCCATTCTG	TTTGTGAATA	TCAAGGCCAA	TCGTCTGACC	TGCTCTAACCC	TCCTGTCAT	2280
2281	GCTGGCGGCG	GCTCTGGTGG	TTGTTCTGTT	GGCGGCTCTG	AGGGTGGTGG	CTCTGAGGGT	2340
2341	GGCGGTTCTG	AGGGTGGCGG	CTCTGAGGG	GGCGGTTCTG	GTGGTGGCTC	TGGTTCGGT	2400
2401	GATTTTGTAT	ATGAAAAGAT	GGCAAAACGCT	AATAAGGGGG	CTATGACCGA	AAATGCGAT	2460
2461	AAAAACCGCG	TACAGTCTGA	CGCTAAAGGC	AAACTTGATT	CTGTCGCTAC	TGATTACGGT	2520
2521	GCTGCTATCG	ATGGTTTCTAT	TGGTGAACGTT	TCCGGCTCTG	CTAATGGTAA	TGGTGTACT	2580
2581	GGTGATTTTG	CTGGCTCTAA	TTCCCTAAATG	GCTCAAGTCG	GTGACGGTGA	TAATTACCT	2640
2641	TTAATGAATA	ATTTCCGTTA	ATATTTACCT	TCCCTCCCTC	AATCGGTTG	ATGTCGCCCT	2700
2701	TTTGTCTTAA	GGCGCTGGTAA	ACCATATGAA	TTTCTATTG	ATTGTCACAA	AATAAACTTA	2760
2761	TTCCGTGGTG	TCTTGTGTT	TCTTTTATAT	GTTGCCACCT	TTATGTTATG	ATTTTCTACG	2820
2821	TTTGCTAACAA	TACTGCGTAA	TAAGGAGTCT	TAATCATGCC	AGTTCTTTG	GGTATCCGT	2880
2881	TATTATTCTG	TTTCTCGGT	TTCTCTCTGG	TAACTTTGTT	CGGCTATCTG	CTTACTTTTC	2940
2941	TTAAAAAGGG	CTTCGGTAAG	ATAGCTATTG	CTATTTCTT	CTTCTTGTGCT	CTTATTATTG	3000
3001	GGCTTAACTC	AATTCTGTG	GGTTTATCTCT	CTGATAATTG	CGCTCAATTAA	CCCTCTGACT	3060
3061	TTGTTCAAGG	TGTTCAAGTTA	ATTCTCCCGT	CTAATGCGCT	TCCCTGTTT	TATGTTATTG	3120
3121	TCTCTGTAAA	GGCTGCTATT	TTCATTTTGT	ACGTTAAACA	AAAAAATCGTT	TCTTATTGTTG	3180
3181	ATTGGGATAA	ATAATATGGC	TGTTTATTTT	GTAACCTGGCA	ATTAGGCTC	TGGAAAGACG	3240
3241	CTCGTTAGCG	TTGGTAAAGG	CTAGGATAAAA	ATTGTTAGCTG	GGTGCAAAT	AGCAACTAAT	3300
3301	CTTGATTTAA	GGCTTCAAAA	CCTCCCGCAGA	GTGGGAGGT	TCGCTAAAAC	GCCTCGCGT	3360
3361	CTTAGAATAC	CGGATAAGCC	TTCTATATCT	GATTGCTTG	CTATTGGGCG	CGGTAATGAT	3420
3421	TCCTACGATG	AAAATAAAAA	CGGCTTGTCTT	GTTCTCGATG	AGTGCCTGAC	TTGGTTTAAT	3480
3481	ACCCGTTCTT	GGAAATGATAA	GGAAAGACAG	CCGATTATTG	ATGGTTCTCT	ACATGCTCGT	3540
3541	AAATTAGGAT	GGGATATTAT	TTTCTTGTGTT	CAGGACTTAT	CTATTGTTGA	TAACACGGCG	3600
3601	CGTTCTGCT	TAGCTGAACA	TGTTGTTAT	TGTCGTCGT	TGGACAGAAAT	TACTTTACCT	3660
3661	TTTGTGCGTA	CTTTATATTG	TCTTATTACT	GGCTCGAAAA	TGCGCTCTGCC	TAATTACAT	3720
3721	GTTGGCGTTG	TTAAATATGG	CGATTCTCAA	TAAAGCCCTA	CTGTTGAGCG	TTGGCTTAT	3780
3781	ACTGGTAAGA	ATTTGTATAA	CGCATATGAT	ACTAAACAGG	CTTTTCTTAG	TAATTATGAT	3840
3841	TCCGGTGT	ATTCTTATT	AACGCCCTAT	TTATCACACG	GTGCGGTATTT	CAAACCATTA	3900
3901	AATTAGGTC	AGAAGATGAA	GCTTACTAAA	ATATATTGAA	AAAAGTTTTC	ACGCCTTCTT	3960
3961	TGTCTTGCAGA	TTGGATTGTC	ATCAGCATT	ACATATAGTT	ATATAACCCA	ACTTAAGCCG	4020
4021	GAGGTTAAA	AGGTAGTCTC	TCAGACCTAT	GATTTGATA	AATTCACTAT	TGACTCTTCT	4080

FIG. 6-1
SUBSTITUTE SHEET

11 / 11

4081	CAGCGTCTTA	ATCTAAGCTA	TCGCTATGTT	TTCAAGGGATT	CTAAAGGGAAA	ATTAATTAAT	4140
4141	AGCGACGATT	TACAGAAGCA	AGGTATTCA	CTCACATATA	TTGATTTATG	TACTGTTTCC	4200
4201	ATTAAAAAG	GTAATTCAA	TGAAATTGTT	AAATGTAATT	AATTTTGTTT	TCTTGAATGTT	4260
4261	TGTTTCATCA	TCTTCCTTGT	CTCAGGTAAT	TGAAATGAAT	AATTGCGCTC	TGCGCGATT	4320
4321	TGTAACCTGG	TATTCAAAGC	AATCAGGCAGA	ATCCGTTATT	GTTTCTCCCAG	ATGTAAAAGG	4380
4381	TACTGTTACT	GTATATTCA	CTGACGTTAA	ACCTGAAAAT	CTACGCAATT	TCTTTATTTTC	4440
4441	TGTTTACGT	GCTAAATAATT	TGATATGGT	TGGTTCAATT	CTCTCCATAA	TTCAGAAAGTA	4500
4501	TAATCCAAAC	AATCAGGATT	ATATTGATGA	ATTGCCATCA	TCTGATAATC	AGGAATATGA	4560
4561	TGATAATTCC	GCTCCCTCTG	GTGGTTTCTT	TGTTCCGCAA	AATGATAATG	TTACTCAAAC	4620
4621	TTTTAAAATT	AATAACGTTC	GGGCAAAGGA	TTAACATACGA	GTGTCGAAT	TGTTTGTAAA	4680
4681	GTCTAACT	TCTAAATCTC	CAAATGTTAT	ATCTATTGAC	GGCTCTAAAT	TATTAGTTGT	4740
4741	TAGTGCAC	AAAGATATT	TAGATAACCT	TCCTCAATT	CTTCTACTG	TTGATTGCCC	4800
4801	AACTGACCG	ATATTGATTG	AGGGTTTGTAT	ATTGAGGTT	CAGCAAGGTG	ATGCTTAGA	4860
4861	TTTTTCAATT	GCTGCTGGCT	CTCAGCGTGG	CACTGTTGCA	GGCGGTGTTA	ATACTGACCG	4920
4921	CCTCACCTCT	GTTTTATCTT	CTGCTGGTGG	TTCGTTGGT	ATTTTTAAATG	GCGATGTTTT	4980
4981	AGGGCTATCA	GTTCGCGCAT	TAAGACTAA	TAGCCATTCA	AAAATATTGT	CTGTGCCACG	5040
5041	TATCTTACG	CTTCAGGT	AGAAGGGTTIC	TATCTCTGTT	GGCCAGAAATG	TCCCTTTTAT	5100
5101	TACTGGTCTG	GTGACTGGTG	AATCTGCCAA	TGTTAAATAAT	CCATTTCAGA	CGATTGAGCG	5160
5161	TCAAAATGTA	GGTATTTCGA	TGAGCGTTTT	TCCTGTTGCA	ATGGCTGGCG	GTAATATTGT	5220
5221	TCTGGATATT	ACCAGCAAGG	CCGATAGTTT	GAGTTCTTCT	ACTCAGGCAA	GTGATGTTAT	5280
5281	TACTAATCAA	AGAAGTATTG	CTACAAACGGT	TAATTGCGT	GATGGACAGA	CTCTTTTACT	5340
5341	CGGTGGCTTC	ACTGATTAA	AAAACACTTC	TCAAGATTCT	GGCGTACCGT	TCCTGCTCAA	5400
5401	AATCCCTTAA	ATCGGCTTC	TGTTTAGCTC	CCGCTCTGAT	TCAACAGGAG	AAAGCACGTT	5460
5461	ATACGTGCTC	GTCAAAGCAA	CCATAGTACG	CGCCCTGTAG	CGCGCATT	AGCGCGGC	5520
5521	GTGTGGTGGT	TACGCGCAGC	GTGACCGCTA	CACTGCCCCAG	CGCCCTAGCG	CCCCTCCCT	5580
5581	TCGCTTTCTT	CCCTCTCTT	CTCGCCACCGT	TCGCCGGCTT	TCCCCGTCAA	GCTCTAAATC	5640
5641	GGGGGCTCC	TTTAGGGTTC	CGATTAGTGT	CTTACGGCA	CCTCGACCCCC	AAAAAAACTTG	5700
5701	ATTTGGGTTGA	TGGTTCACGT	AGTGGGCCCCAT	CGCCCTGATA	GACGGTTTTT	CGCCCTTTGA	5760
5761	CGTTGGAGTC	CACGTTCTT	AAATGTTGGAC	TCTTGTCCA	AACTGGAACA	ACACTCAACC	5820
5821	CTATCTCGGG	CTATTCTTT	GATTTATAAG	GGATTTTGC	GATTTTGGAA	CCACCATCAA	5880
5881	ACAGGATT	CGCCCTGCTGG	GGCAGAACCG	CGTGGACGCC	TTGCTGCAAC	TCTCTCAGGG	5940
5941	CCAGGGGGT	AAGGGCAATC	AGCTGTTGCG	CGTCTCGCTG	GTGAAAAGAA	AAACCACCC	6000
6001	GGCGCCCAAT	ACGCAAAACCG	CCTCTCCCCG	CGCGTTGGCC	GATTCAATTAA	TGAGCTGGC	6060
6061	ACGACAGGGT	TCCCGACTGG	AAAGCGGGCA	GTGAGCGCAA	CGCAATTAAAT	GTGAGTTAGC	6120
6121	TCACTCATTA	GGCACCCCCAG	GCTTTACACT	TTATGTTTCC	GGCTCGTATG	TTGTGTGGAA	6180
6181	TTGTGAGCGG	ATAAACATT	TACACGCCAA	GGAGACAGTC	ATAATGAAAT	ACCTATTGCC	6240
6241	TACGGCAGCC	GCTGGATTGT	TATTAATCTC	TGCCCCAACCA	GCATGGCCG	AGCTCTTCCC	6300
6301	GCCATCTGAT	GAGCAAGTGA	AATCTGGAAC	TGCTCTGTT	GTGTGCTGC	TGAATAACTT	6360
6361	CTATCCCAGA	GAGGCCAAAG	TACAGTGGAA	GGTGGATAAC	GCCCTCCAAT	CGGGTAACTC	6420
6421	CCAGGAGAGT	GTCACAGAGC	AGGACAGCAA	GGACAGCACC	TACAGCTCTA	GCAGCACCC	6480
6481	GACGCTGAGC	AAAGCAGACT	ACGAGAACAA	CAAAGCTCTA	GCCTGCGAAG	TCACCCATCA	6540
6541	GGGCCTGAGC	TGCCCCGTCA	AAAGAGCTT	CAACAGGGGA	GAGTGTCTA	GAACGGTCA	6600
6601	CTTGGCACTG	GCCGTCGTTT	TACAACGTCG	TGACTGGGAA	AACCCCTGGCG	TTACCCAAAGC	6660
6661	TTTGTACATG	GAGAAAATAA	AGTGAACAA	AGCACTATTG	CACTGGCACT	CTTACCGTT	6720
6721	CTGTTTACCC	CTGTGGCAAA	AGCCGCTCC	ACCAAGGGCC	CATCGGTCT	CCCCCTGGCA	6780
6781	CCCTCTCTCA	AGAGCACCTC	TGGGGGCA	GGGGCCCTGG	GCTGCTGGT	CAAGACTAAT	6840
6841	CCCCCGAAC	GGTGAAGGTG	TCGTGGAACT	CAGGGCCCT	GACCAGCGGC	GTGCACACCT	6900
6901	TCCCGGTGT	CCTACAGTCC	TCAGGACTCT	ACTCCCTCAG	CAGCGTGGTG	ACCGTGCCCT	6960
6961	CCAGCAGCTT	GGGCACCCAG	ACCTACATCT	GCAACGTGAA	TCACAAGCCC	AGCAACACCA	7020
7021	AGGTGGACAA	GAAAGCAGAG	CCAAATCTT	GTACTAGTGA	ATCTTACCCG	TACGACGTT	7080
7081	CGGACTAGC	TTCTTAGGCT	AAAGGGCATG	ACCCCTGCTAA	GGCTGCACTT	AATAGTTTAC	7140
7141	AGGCAAGTGC	TACTGAGTAC	ATTGGCTACG	CTTGGGCTAT	GGTAGTAGTT	ATAGTTGGTG	7200
7201	CTACCATAGG	GATAAAATT	TTCAAAAAGT	TTACGAGCAA	GGCTTCTTAA	GCAATAGCGA	7260
7261	AGAGGCCCGC	ACCGATCGCC	CTTCCCAACA	GTTGCGCAGC	CTGAATGGCG	AATGGCGCTT	7320
7321	TGCTGGTTT	CCGGCACCAAG	AAGCGGTGCG	GGAAAGCTGG	TTGGAGTGC	ATCTTCCCTGA	7380
7381	GGCCGATACG	GTCGCTGTC	CCTCAAACTG	GCAGATGCA	GTTTACGATG	CGCCCCATCTA	7440
7441	ACCAACAGTA	ACCTATCCC	TTACGGTCAA	TCCGCCGTTT	GTTCCCACGG	AGAATCCGAC	7500
7501	GGGTTGTTAC	TCGCTCACAT	TTAATGTTGA	TGAAAGCTGG	CTACAGGAAG	GCCAGACGCG	7560
7561	AATTATTTT	GATGGCGTT	CTATTGGTTA	AAAAATGAGC	TGATTTAAC	AAAATTTAAC	7620
7621	GCGAATT	ACAAATATT	AACGTTTAC	ATTTAAAT	TTGCTTATAC	AATCTTCTG	7680
7681	TTTTGGGGC	TTTCTGTT	ATCAACCGGG	GTACATATGA	TTGACATGCT	AGTTTACGA	7740
7741	TTACCGTCA	TCGATTCCT	TGTTTGCCT	AGACTCTCAG	GCAATGACCT	GATAGCTTT	7800
7801	GTAGATCTC	AAAAAATAGC	TACCCCTCTCC	GGCATTAAATT	TATCAGCTAG	AACGGTTGAA	7860
7861	TATCATATTG	ATGGTGTATT	GACTGTCCTC	GGCCTTCTC	ACCCCTTTGA	ATCTTACCT	7920
7921	ACACATTACT	CAGGCATTGC	ATTTAAAATA	TATGAGGGTT	CTAAAAAATTT	TTATCCTTGC	7980
7981	GTTGAAATAA	AGGCTTCTCC	CGCAAAAGTA	TTACAGGGTC	ATAATGTTT	TGGTACAACACC	8040
8041	GATTTAGCTT	TATGCTCTGA	GGCTTATTG	CTTAATTTG	CTAATTCCTT	GCCTTGCCCTG	8100
8101	TATGATTAT	TGGACGTT					8118

FIG. 6-2
SUBSTITUTE SHEET

INTERNATIONAL SEARCH REPORT

International Application No.

PCT/US91/07149

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all) ⁶

According to International Patent Classification (IPC) or to both National Classification and IPC
 IPC(5): C12N 15/64, 15/70
 U.S.C1.: 435/252.3, 320.1

II. FIELDS SEARCHED

Minimum Documentation Searched ⁷

Classification System	Classification Symbols
U.S.C1.	435/69.7, 172.3, 252.3, 320.1

Documentation Searched other than Minimum Documentation
to the Extent that such Documents are Included in the Fields Searched ⁸

APS, STN/MEDLINE, TERMS USED: SURFACE EXPRESSION VECTOR#, DIRECTED EVOLUTION, SINGLE CHAIN ANTIBOD?.

III. DOCUMENTS CONSIDERED TO BE RELEVANT ⁹

Category ¹⁰	Citation of Document, ¹¹ with indication, where appropriate, of the relevant passages ¹²	Relevant to Claim No. ¹³
Y	W.C.A., 28/CC630 (FOM ET AL) 07 September 1989 see entire document.	1-75
Y	Nucleic Acids Research, Vol. 12, No. 9, issued SEPTEMBER 1984, BCSS ET AL, "Assembly of functional antibodies from immunoglobulin heavy and light chains synthesized in <i>E. coli</i> ", pages 3731-3806, see the abstract.	5-75
Y	Proceedings of the National Academy of Sciences, Vol. 86, issued AUGUST 1989, SASTRY ET AL, "Cloning of the immunological repertoire in <i>Escherichia coli</i> for generation of monoclonal catalytic antibodies: Construction of a heavy chain variable-region specific cDNA library", pages 5728-5732, see the abstract.	1-75
Y	Science, Vol 246, issued 08 December 1989, Huse et al, "Generation of a Large Combinatorial Library of the Immunoglobulin Repertoire in Phage Lambda", pages 1275-1281, see entire document.	1-75

* Special categories of cited documents: ¹⁰

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"G" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search
06 January 1992

Date of Mailing of this International Search Report

21 JAN 1992

International Searching Authority
ISA/US

Signature of Authorized Officer
John D. Ulm
John D. Ulm

FURTHER INFORMATION CONTINUED FROM THE SECOND SHEET

Y

Gene, Vol. 73, issued 1998, PARMLEY ET AL.
"Antibody-selectable filamentous fd phage
vectors: affinity purification of target
genes", pages 305-318, see entire document.

6-75

V. OBSERVATIONS WHERE CERTAIN CLAIMS WERE FOUND UNSEARCHABLE¹

This international search report has not been established in respect of certain claims under Article 17(2) (a) for the following reasons:

1. Claim numbers . . . because they relate to subject matter¹² not required to be searched by this Authority, namely:

2. Claim numbers . . . because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out¹³, specifically:

3. Claim numbers _____, because they are dependent claims not drafted in accordance with the second and third sentences of PCT Rule 6.4(a).

VI. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING²

This International Searching Authority found multiple inventions in this international application as follows:

1. As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims of the international application.

2. As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims of the international application for which fees were paid, specifically claims:

3. No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claim numbers:

4. As all searchable claims could be searched without effort justifying an additional fee, the International Searching Authority did not invite payment of any additional fee.

Remark on Protest

- The additional search fees were accompanied by applicant's protest.
 No protest accompanied the payment of additional search fees.